# Population genetics of native and domesticated *Eucalyptus globulus*

Tim H. Jones

B. Sc. (Hons)

Submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy

School of Plant Science, University of Tasmania
June, 2005

# Declarations

This thesis does not contain any material which has been accepted for a degree or diploma by the University of Tasmania or any other institution. To the best of my knowledge and belief this thesis contains no material previously published or written by another person, except where due acknowledgement is made in the text.

This thesis may be made available for loan and limited copying in accordance with the Copyright Act 1968.

Signed: Date: 8/11/2005

Tim H. Jones

# Abstract

This thesis consists of firstly, a study of the local genetic dynamics occurring in a native population of *Eucalyptus globulus*, and secondly, an examination of genetic diversity within the Australian *E. globulus* breeding population.

Fine scale spatial genetic structure within a continuous *E. globulus* forest on the Tinderbox Hills, Tasmania, was examined by genotyping mature trees (n = 168) and juveniles (n = 110) in a 140 m diameter sample site, using microsatellite markers. Spatial genetic autocorrelation revealed a significant decline in genetic similarity with distance between individuals, with the mature cohort displaying greater spatial genetic structure than the juvenile cohort. High-resolution analysis, using a combination of Bayesian clustering, ordination and spatial interpolation, revealed a complex pattern of genetic groups superimposed upon each other. Comparison between cohorts, and parentage analysis, indicated a directional shift in the distribution of genetic variation over generations, despite no differences in overall genetic diversity being detected between cohorts.

Paternity analysis of 549 open-pollinated seed from nine trees identified 374 mating pairs. Sixty-seven percent of the pollinations involved trees from within the sample population where pollen dispersal was leptokurtic but not directional. Threshold relatedness values for full-sibling, half-sibling and unrelated relationships were calculated by simulation to quantify bi-parental inbreeding. On average, 10% of progeny were derived from self-pollination, with a further one percent and 13% derived from crossing between likely full- and half-siblings respectively. A further 43% of progeny were the product of unrelated individuals within the study site. Significant differences between families in the level of self-pollination and bi-parental inbreeding were revealed. Seed, juvenile and mature cohorts of the population showed high levels of observed and expected heterozygosity and there was little evidence for selection against homozygotes across cohorts.

Genetic diversity within the first generation of the Australian *E. globulus* breeding population was compared with that observed between mature trees sampled from the entire natural distribution of the species. Observed and expected heterozygosity in the breeding populations was at least as high as that displayed by the native samples, and there was no evidence of increased inbreeding in the breeding population. The breeding population captured a significant amount of genetic variation from the natural distribution of the species with most selections fitting closely to their native race of origin. However, a number of pedigree errors were detected at the level of race, family and genotype. This study provides a benchmark for monitoring genetic diversity during the course of domestication of *E. globulus*.

# Acknowledgements

This project would not have been possible without the help of many people. Firstly, I would like to thank my supervisors René Vaillancourt and Brad Potts for their dedicated and enthusiastic support. The passionate discussions and late night sessions were an integral part of my learning over the course of this project. I greatfully acknowledge the support of the Australian Research Council and the Southern Tree Breeders Association (STBA) for their financial support of this project. I particularly thank Francis Cole for allowing me to study the forest on his land, and the Kingborough Council for assisting in access to this land.

Many thanks are due to Adam Smolenski for his invaluable help in the lab, Beck Jones for her significant assistance with genotyping and Dot Steane for allowing me to use her microsatellite dataset of the native distribution of *E. globulus* in this study. I particularly thank Paul Tilyard for his years of technical support and insightful advice. I am grateful for the advice and friendship provided by Carl Grosser who introduced me to a number of essential analytical tools. Bruce Greaves generously gave a significant amount of his time and I am indebted to Sophie Gerber for her advice and proof reading.

I am fortunate to have been surrounded by a number of friendly and helpful people during the course of this project. Peter Gore, Greg Dutkowski, and Greg Jordon provided significant advice and amusement over the years. I thank those I have shared office and lab space with for their support, humour and tolerance. In particular, thanks to Gay McKinnon for her friendship, advice (both scientific and artistic), proof reading and risotto. It was a pleasure to share my office with Jules Freeman.

Finally many thanks to my friends and family. My parents, Ross and Jan Jones provided significant financial and emotional support during this project. Thank you to my partner Sophie Le Roux for her love, patience, support and proof reading.

# Table of Contents

# Chapter 1: General Introduction

## Background

The essence of Darwinian evolutionary theory is the conversion of genetic variability between interbreeding individuals into genetic variation between groups of individuals over time and space, eventuating in speciation (Lewontin 1974). Gene flow, selection and genetic drift cause genetic mutations at the individual level to permeate every strata of the hierarchy of life (Kimura and Ohta 1971; Hartl and Clark 1989). In forest species, the ongoing interplay between these processes creates a dynamic continuum of genetic diversity that is distributed both temporally and spatially throughout populations and across the landscape (Wright 1951; Ehrlich and Raven 1969; Crow 1986). The adaptive potential of a forest species is dependent on the distribution of genetic diversity within and between populations, the mating system of the species and how these variables interact and perpetuate each other (Hamrick 1983; Epperson 1992).

## Molecular markers in population genetics

Before the development of molecular markers, the study of genetic variation within species and mating systems was reliant on the use of quantitative genetic techniques to separate genetic from environmental effects on phenotype (Fins *et al.* 1992; Wang and Szmidt 2001). While applicable to plant and animal breeding where specific traits are of interest and pedigree history is well known, classical quantitative genetics based on phenotype was limited in its ability to unravel the complex genetic relationships involved in population dynamics (Lewontin 1985). The development and application of molecular techniques to population genetics during the 1980's provided a method of characterizing individuals based on heritable and selectively neutral traits (Sunnocks 2000). This resulted in a revolution in various aspects of population genetics and evolutionary studies, including studies of gene flow, mating systems, genetic drift, and population genetic diversity (Avise 1994; Parker 1998; Davies *et al.* 1999; Wang and Szmidt 2001; Lewontin 2002).

Since the initial development of protein electrophoresis, molecular techniques have rapidly advanced to the point where genetic variation between individuals, populations, races and species is commonly characterized with high definition (Wang and Szmidt 2001). Different regions of DNA experience different rates of mutation over time (Wolfe *et al.* 1987; Wolfe *et al.* 1989; Zhang and Hewitt 2003) and this phenomenon has facilitated the development of molecular markers that display various rates of change (Avise 1994; Zhang and Hewitt 2003). The nuclear, mitochondrial and chloroplast genomes within plants are characterized by different organization, inheritance, rates and patterns of evolution, with DNA markers derived from each displaying specific applications to the various levels of genetic hierarchy of interest in population genetics (Avise 1994; Neigel 1997). Hence, molecular markers now allow information to be gained from all levels of population and evolutionary process with a high level of resolution (Sunnocks 2000). Recent advances in genotyping techniques now allow the rapid accumulation of large amounts of highly informative genetic data (Glaubitz and Moran 2000). In addition, the development of computational resources, software and statistical techniques has provided a crucial platform for complex analysis of large amounts of molecular data (Pritchard *et al.* 2000a; Jones and Ardren 2003; Manel *et al.* 2003; Holsinger and Wallace 2004; Smouse and Sork 2004).

Allozymes were the first molecular markers to be applied to population genetics on a wide scale (Hamrick and Godt 1990; Parker 1998). These markers are polymorphic forms of inherited enzymes that are separated using gel electrophoresis and used to categorize individuals and differentiate between various levels of genetic grouping (e.g. populations, races or species) (Parker 1998). Allozyme markers are still commonly used at present, primarily due to a codominant mode of inheritance, the applicability to virtually all species, a relatively inexpensive cost and technical simplicity (Ouborg *et al.* 1999). However, the coding region of DNA represents a small fraction of the total DNA and hence, the proportion of DNA polymorphism that is detectable with allozyme markers is restricted (Avise 1994). In comparison to more recently developed markers, allozymes typically display relatively few polymorphisms, and polymorphisms may also be quite difficult to resolve using gel electrophoresis, resulting in a substantial amount of variation between alleles remaining undetected (Avise 1994). In many cases, silent

DNA mutations may not result in a detectable change in the corresponding allozyme (Avise 1994). The presence of allozyme markers is reliant on the particular protein being expressed and so dependent to some extent on regulatory processes and environmental effects (Poly 1997).

DNA markers, on the other hand, are completely heritable and are therefore the most accurate source of information for the characterization of genetic relationships between individuals (Avise 1994; Parker 1998). While generally more expensive to use than allozyme markers, DNA markers provide an infinitely higher resolution of genetic characterization, to the point where sequencing of DNA regions, or indeed the complete genome, allows the identification of single nucleotide polymorphisms (SNP's) between individuals (Avise 1994; Vignal *et al.* 2002). Focusing on the coding region of DNA can provide detailed information about the heritability of adaptive traits, locate quantitative trait loci, and identify the functional candidate genes for the traits in question (Ruiz and Barbadilla 1995; Pflieger *et al.* 2001; Rieseberg and Burke 2001). On the other hand, analysis of the non-coding region of the chromosome can provide potentially selectively neutral information that is required for the study of population genetic dynamics (Hartl and Clark 1989). In addition, comparison of the variation in quantitative traits and neutral genetic markers can provide an estimate of adaptive population divergence (McKay and Latta 2002; Latta 2004).

**Microsatellite markers**

At present, a wide range of DNA markers are available for use in population genetics (e.g. microsatellites (SSR's or STR's), RFLP's, RAPD's, AFLP's, VNTR's and SNP's) and these are reviewed in numerous publications including Avise (1994), Hartl and Clark (1989), Parker (1998), Vignal *et al.* (2002), and Wang and Szmidt (2000). Over the last 10 years, microsatellite loci have emerged as extremely popular markers for population genetics studies due particularly to their high mutation rate, their consequent high level of polymorphism, high frequency in many organisms, their codominant mode of inheritance, and the relative ease of fragment length scoring (Jarne and Lagoda 1996; Lefort 1999). Microsatellite loci are characterized by tandemly repeated sequences, with

between one and five (usually two) nucleotides being repeated (Jurka and Pethiyagoda 1995). Once identified, microsatellite regions are amplified with primers that are specific to the flanking regions of the repeated sequence, using the polymerase chain reaction (PCR) (Avise 1994; Zane *et al.* 2002). The difficulty in development of the often species specific microsatellite primers (Zane *et al.* 2002) and potential for homoplasy when comparing more distant lineages (Estoup *et al.* 2002) should not be left ignored. However, microsatellite loci are now routinely used in many aspects of population genetic analysis and plant breeding (Goldstein and Schlötterer 1999), including parentage analysis (e.g. Jones and Ardren 2003, Gerber *et al.* 2000), kinship and ancestry studies (e.g. Queller *et al.* 1993, Van de Casteele 2001, Blouin *et al.* 1996 and Rosenberg *et al.* 2003), gene flow studies (e.g. Austerlitz *et al.* 2004, Ouborg *et al* 1999 and Neighel 1997), spatial genetic analysis (e.g. Pearse and Crandall 2004, Smouse and Peakall 1999, Beaumont 1999 and Pritchard *et al.* 2000), association mapping and quantitative trait loci analysis (e.g. Pritchard *et al.* 2000b, Shepherd and Jones 2005 and Wang and Szmidt 2001), quality control (e.g. Lexer *et al.* 2001 and Nybom 1991), analysis of genetic diversity (e.g. Shepherd and Jones 2005, Glaubitz and Moran 2000 and Jarne and Lagoda 1996) and many more.

**Thesis outline**

In the current study microsatellite markers were used to examine a range of questions regarding the population genetics of the Tasmanian blue gum *Eucalyptus globulus* (*sensu* Brooker 2000) both in its natural environment and in the Australian *E. globulus* breeding program. At present a range of microsatellite loci have been developed that are directly applicable to *Eucalyptus globulus* (Brondani *et al.* 1998; Steane *et al.* 2000). A number of these loci are highly polymorphic within the species, providing an excellent tool for revealing many aspects of the complex population genetic dynamics of the species. The specific aims of this thesis are outlined below:

1. To quantify and characterize local spatial genetic variation in a continuous native population of the species.

2. To compare spatial genetic variation between two life stages of the species in its native environment.
3. To quantify the amount of bi-parental inbreeding that naturally occurs within the species.
4. To investigate localized patterns of gene flow in the species, in its native environment.
5. To quantify the amount of genetic diversity within the Australian *E. globulus* breeding program.
6. To carry out quality control testing of pedigrees within the Australian *E. globulus* breeding program.

These investigations will provide insight into historical and on-going genetic dynamics of *E. globulus* within natural forest and domesticated systems. In addition to further describing processes that underlie the evolution of forest species, the quantification of spatio-temporal genetic patterns, gene flow and how these variables interact with each other will contribute information that is directly relevant to the future management and conservation of forest species in the natural environment. As natural forest resources continue to decline, tree breeding programs of domesticated forest species may also be optimized to assist in the conservation of genetic diversity of a species (Lefèvre 2004). Hence, pedigree verification and assessment of genetic diversity within the Australian *E. globulus* breeding program is a major step in providing a benchmark for the management of genetic diversity of the species into the future.

**Structure of the thesis**

The three experimental chapters (Chapter 2 to 4) are presented as self contained units, including in-depth introduction to the techniques used and discussion of the biological implications of the studies. Chapter 2 examines current patterns of spatial genetic variation in two cohorts of a continuous population of *E. globulus*. In addition to addressing the biological questions raised in this chapter, the discussion of current spatial genetic analysis techniques has also been emphasized. Chapter 3 provides the

necessary investigation into the genetic dynamics that are influencing (and influenced by) the spatial genetic structure revealed in Chapter 2. Genetic diversity within three life stages of the species in the natural environment is also compared in Chapter 3. In Chapter 4, the focus shifts from the localized single population studies of the first two chapters to genetic diversity on a species wide scale. Genetic diversity within the Australian national *E. globulus* breeding program is assessed, in comparison to that evident throughout the native distribution of the species. In addition, a quality control analysis of the pedigrees within the breeding program is included in this chapter. Due to the level of discussion within each methodological chapter, Chapter 5 provides a brief thesis discussion that summarizes the most important findings of the thesis.

# Chapter 2: Spatial genetic structure in *Eucalyptus globulus*

## Introduction

Within forest populations, a number of demographic processes tend to induce various strata of spatial genetic structure (Epperson 2000). These include micro-habitat variation in selection pressure (Bradshaw 1972; Endler 1973; Slatkin and Arter 1991), spatial distribution and density of individuals (Levin 1974; Hamrick and Nason 1996; Doligez and Joly 1997) and both spatial and temporal patterns of regeneration (Levin 1988; McCauley *et al.* 1988; Wade and Mccauley 1988; Schnabel and Hamrick 1995). However, even in the absence of these factors, limited gene flow within a population is sufficient alone to produce a mosaic of genetically congruent patches of individuals, as described by the theory of isolation by distance (Wright 1943; Linhart 1989; Epperson 1992). Under conditions of limited pollen or seed dispersal, spatial genetic structure may be particularly influenced by local pedigree structure and various levels of proximity based consanguineous breeding, resulting in a relationship between kinship and proximity at a fine geographic scale (Malécot 1948; Turner *et al.* 1982; Epperson 1995a; Vekemans and Hardy 2004).

In recent years, spatial genetic structure in plant populations has been predominantly described using various forms of spatial autocorrelation analysis (Sokal and Oden 1978; Sokal *et al.* 1989; Legendre 1993) based on data obtained from an assortment of molecular markers (Heywood 1991; Epperson 2000; Wang and Szmidt 2001). This method describes the average genetic distance between individuals according to geographic distance (Sokal and Oden 1978). Spatial autocorrelation analysis is routinely applied to native forest populations (Epperson 1992; Leonardi and Menozzi 1996; Le Corre *et al.* 1998; Streiff *et al.* 1998; Latouche-Halle *et al.* 2003) and this approach has become particularly powerful with the recent availability of high throughput genotyping techniques (Wang and Szmidt 2001) and increased computational resources (Schnabel *et al.* 1998). However, spatial genetic autocorrelation analysis is dependent on the method by which genetic similarity between individuals is estimated (Epperson 1995a) and is limited in its descriptive power (Vekemans and Hardy 2004). Since the development of

joint-count statistics (Epperson 1995b) and the widely used Moran's *I* statistics (Moran 1950), a number of genetic distance measures have been developed and used in spatial genetic analysis (Heywood 1991; Smouse and Peakall 1999; Epperson 2000; Vekemans and Hardy 2004). A significant proportion of these attempt to specifically quantify the degree of pedigreed relatedness (also referred to as kinship or as a probability of identity by descent) between individuals based on putatively neutral markers (Queller and Goodnight 1989; Li *et al.* 1993; Loiselle *et al.* 1995; Ritland 1996; Lynch and Ritland 1999; Wang 2002). The ability of these various relatedness coefficients (Blouin 2003) to directly indicate a specific level of genealogical relationship in spatial genetic analysis is a significant advantage over the traditional genetic distance measures from which quantification of spatial genetic association in specific terms of pedigreed relatedness is much more difficult to elucidate (Vekemans and Hardy 2004).

When variation between spatially discrete units such as populations, or groups within populations is of interest, a common approach is to apply a group-wise genetic distance measure such as Wright's *F*-statistics (Wright 1943) or Nei's genetic distance (Nei 1972; Nei 1978). A wide range of genetic distance measures (either independently derived or variations on the classic measures) have been developed for specific molecular markers and specific genetic questions (Goldstein *et al.* 1995; Shriver *et al.* 1995; Slatkin 1995; Kalinowski 2002), however *F*-statistics are most commonly used (Excoffier 2001; Weir and Hill 2002; Balding 2003). Similarly, a wide range of methods may be used to test the significance of genetic variation described by these various distance measures (Holsinger 1999; Ryman and Jorde 2001; Weir and Hill 2002). Alternatively, Paetkau *et al.* (1995) have developed an assignment test to evaluate the degree of genetic differentiation (and gene flow) between populations. The observed allele frequencies in each population are used as a reference to assign individual genotypes, based on the likelihood of that genotype belonging to a particular gene frequency combination (Paetkau *et al.* 1998; Kyle *et al.* 2000).

Group-wise comparisons of genetic distance based on summary statistics such as $F_{st}$ or assignment tests depend on predefined discrete populations or groups within populations (Hardy *et al.* 2000). As pointed out by Pritchard *et al.* (2000), the definition of populations and other genetically discrete groups is typically subjective and is usually

determined by some tangible factor such as variation in geography, environment or morphology. Continuous forest populations may display "cryptic" genetic structure where genetic structure may exist but be impossible to detect using visible characters (Pritchard *et al.* 2000a) and the formation of suitable testable hypothesis regarding fine-scale genetic subdivisions may be extremely difficult. The recent application of Bayesian statistical methods to the study of population structure has presented a solution to this problem by providing a framework for clustering individuals into genetically similar groups, without the need to specifically predefine the boundaries of these groups (Pritchard *et al.* 2000a; Rosenberg *et al.* 2002). Bayesian methods have also been applied to the verification of genetic differentiation among populations by testing estimates of $F_{is}$ and $F_{st}$ (Holsinger 1999; Holsinger *et al.* 2002; Holsinger and Wallace 2004), assignment of individuals to specific populations and quantification of dispersal and migration (Vazquez-Dominguez *et al.* 2001; Cegelski *et al.* 2003; Wilson and Rannala 2003; Berry *et al.* 2004; Lucchini *et al.* 2004), forensics (Foreman *et al.* 1997; Manel *et al.* 2002), inferring common ancestry and detailed pedigree analysis (Rosenberg *et al.* 2001; Rosenberg *et al.* 2003; Parker *et al.* 2004), and association mapping in genetically structured populations (Pritchard and Rosenberg 1999; Pritchard *et al.* 2000b; Ardlie *et al.* 2002). Having defined genetic groups with this model-based approach, individuals can be probabilistically assigned (or re-assigned) to these groups, taking into account the potential for admixture (Pritchard *et al.* 2002). Having defined groups based purely on genotype, individuals from each group can be geographically mapped and the geographic relationships between groups visualized (Cegelski *et al.* 2003). This method has a number of major advantages over more traditional approaches. Firstly, it can independently define specific genetic sub-divisions if they exist and, as it does not rely on global spatial genetic correlations, it may be able to successfully display spatial genetic patchiness under conditions where global spatial autocorrelation and other techniques may not. These conditions include non-uniform distributions of genetic variation, genetic variation arranged in a 'checkerboard distribution' and spatial superimposition or overlap of genetically distinct groups (e.g. Doligez *et al.* 1998). Secondly, the purely genetic grouping is based on multidimensional data and it does not require prior summarization of genetic distance

with the use of a traditional genetic distance coefficient which will depend on various assumptions (Pritchard *et al.* 2000a).

One disadvantage of grouping methods, such as the approach previously presented, is that distinct specific genetic boundaries between groups of individuals are inferred. This is however, rarely the case in continuous forest (Endler 1973), with overlap of family groups and clinal variation more likely to produce general trends in genetic variation (and varying levels of admixture) related to space (Endler 1977; Futuyama 1998). An alternative way to visually summarize geographic patterns of genetic variation is to plot a smoothed genetic distance measure in the form of contours (isopleths) against two-dimensional spatial position. Smoothing of data summarizes trends from exact data points (i.e. individual trees) into a continuous surface that best fits the data (Menozzi *et al.* 1978; Piazza *et al.* 1981). This 'synthetic mapping' of gene frequencies has been extensively applied in the study of human genetic structure and migration patterns (Balanovskaya and Nurbaev 1997; Malaspina *et al.* 1998; Cavalli-Sforza 2001; Fagundes *et al.* 2002; Di Giacomo *et al.* 2003; Pavesi 2004). Commonly, ordination analysis is carried out on genetic data to summarize the multidimensional association between individuals into several informative independent dimensions that, in turn, represent the major directions of genetic variation within the population (Cavalli-Sforza 1966; Sokal *et al.* 1989; Pavesi 2004). In this case, these principal coordinates (or principal components, depending on the type of analysis) (Shi 1993) are representative of an informative multilocus pattern of allele frequencies or allele frequency combinations (Cavalli-Sforza 2001) . Thus, the relative position of individuals along the particular principal coordinate axis can be regarded as an estimate of their similarity due to variation in a particular combination of multilocus allele frequencies. When individual values of descriptive principal coordinates axes within a population are smoothed and geographically plotted (Menozzi *et al.* 1978; Hanotte *et al.* 2002; Pavesi 2004) the result is an effective visual representation of different aspects of genetic structure within a population. This, in turn, may represent different directions of gene flow through a population (Piazza *et al.* 1981; Hanotte *et al.* 2002).

When ordination analysis is combined with spatial autocorrelation analysis, the result is a correlation between geographic distance and the principal coordinate value (Piazza *et*

*al.* 1981). In this case, the principal coordinate used can be regarded a specific measure of one direction in genetic variation within the population. This approach provides a global summary of the spatial association of the particular principal coordinate within the population. It is graphically represented and a statistical test of the significance of that direction in genetic variation can be applied (Oden 1984). Due to the relationship between spatial genetic structure within a population and gene flow (Sokal and Oden 1977; Epperson 2000), a particular geographic association with genetic distance (as is displayed by the plot of the autocorrelation coefficient in this case) may be diagnostic of a specific historic or present mechanism of gene flow (Epperson 1992; Sokal *et al.* 1997; Hardy and Vekemans 1999; Dick *et al.* 2003; Fenster *et al.* 2003; Austerlitz *et al.* 2004). For example, spatial genetic structure due to limited seed dispersal and proximity based inbreeding is likely to display a rapid, leptokurtic decline in spatial genetic autocorrelation (Epperson 1995a; Epperson 1995b). However, a gradual change in selection pressure throughout a population may result in a more gradual, clinal decay in the association between spatial and genetic distance (May *et al.* 1975; Sokal *et al.* 1997). Hence, spatial autocorrelation of various principal coordinate axes may provide an effective method of summarizing and distinguishing between various strata of spatial genetic structure within a population (Sokal *et al.* 1997; Epperson 2000).

Spatial genetic structure within native populations is one of the dominant issues in population genetics (Epperson and Li 1996), influencing (and responding to) a wide range of factors, specifically mating system, inbreeding, and response to selection (Endler 1977; Hartl and Clark 1989; Epperson 1995a). It may illustrate fundamental evolutionary mechanisms and provide evidence for historical processes within a population. From a conservation point of view, spatial structure and associated isolation by distance may inhibit unrelated out-crossing, leading to genetically depauperate populations that may suffer from inbreeding depression, or at least, have a reduced ability to respond to change in selection pressure (Ellstrand 1992b). In addition, spatial genetic structure will affect estimates of population genetic diversity (Wahlund's principal) and hence, its quantification allows increased accuracy in these estimations (Hartl and Clark 1989). Quantification of spatial genetic structure in native forest populations is of importance to breeding programs that originate from native open

pollinated seed. Inbreeding depression arising from crossing related individuals may be avoided, or at least accounted for if this factor is accurately assessed (Borralho and Potts 1996).

*Eucalyptus globulus* Labill. is a widespread forest tree species, naturally distributed throughout Tasmania and the south-east of mainland Australia (Williams and Potts 1996). The species is of significant commercial importance, with extensive plantations established in temperate regions throughout the world, including Australia, Chile, Portugal and Spain (Eldridge *et al.* 1993). In its native range, it is a subdominant and occasional dominant tree of both wet and dry sclerophyll forest (Williams and Potts 1996). As with most eucalypts (e.g. *E. obliqua*, *E. citriadora*, *E. rhodantha*, *E. pauciflora* and *E. kitsoniana*, Potts and Wiltshire 1997) it displays a mixed mating system (Hardner 1996; Potts and Wiltshire 1997; Patterson *et al.* 2001; Patterson *et al.* 2004), exhibiting varying levels of self incompatibility (Hardner *et al.* 1996; Patterson *et al.* 2000; Pound *et al.* 2003). In fact, the level of self pollination is specific to the position of flowers within the canopy, with the level of out crossing increasing with the height of flowers within the canopy (Patterson *et al.* 2001; Patterson *et al.* 2004). The species is pollinated by a range of bird and insect visitors (Hingston and Potts 1998a; Hingston 2002; Hingston *et al.* 2004a; Hingston *et al.* 2004b). The seeds of *E. globulus* are relatively heavy and do not have any specialized dispersal mechanism (Boland *et al.* 1980), indicating that seed dispersal is likely to be limited to a radius of around twice the tree height (Cremer 1966; Cremer 1977). Limited seed dispersal has indeed been suggested to occur in *E. globulus* (Potts and Wiltshire 1997; Skabo *et al.* 1998) among other eucalypts (e.g. *E. rhodantha*, Sampson *et al.* 1989, *E. delegatensis*, Moran and Griffin 1983, *E. amygdalina* and *E. risdonii* Potts and Wiltshire 1997) and is expected to result in a mosaic of related groups of individuals that share relatively common ancestry distributed throughout eucalypt forest (Eldridge *et al.* 1993). In a study of fragmented *E. globulus* populations, using randomly amplified polymorphic DNA (RAPD) markers, Skabo *et al.* (1998) showed that genetically similar *E. globulus* trees occurred within 25 m of each other, however the spatial autocorrelation analysis did not identify any further spatial genetic association beyond this distance class. This was consistent with the spatial structure suggested to occur in *E. globulus* forest by Hardner *et al.* (1998), who

used progeny vigor as an indirect measure of parental relatedness, suggesting that relatedness between mature trees did not further decline beyond 50 m.

With the indication that spatial structure does occur in *E. globulus*, a continuous native pure stand population of the species was studied south of Hobart, Tasmania, Australia. Using eight highly polymorphic microsatellite loci, a population sample of 168 mature trees and 110 juvenile seedlings (all individuals) within a radius of 70 m was genotyped. The specific aims of this study where: to verify the existence of spatial structure in a continuous native *E. globulus* forest using a large number of individuals; to dissect and describe the actual spatial state of genetic structure within the study population; to compare the spatial genetic structure between mature and juvenile cohorts within this population, with a view to gaining insight into the potential change in spatial genetic structure over time. Due to the specificity of the various methods of analysis reviewed in this introduction, an integrated approach was employed in this study, utilizing the positive aspects of spatial autocorrelation analysis, classical population based genetic statistics, Bayesian clustering and assignment, ordination and spatial interpolation to quantify and visualize a complex pattern of fine scale spatial genetic structure.

## Materials and Methods

### Plant Material

A native population of *Eucalyptus globulus* was studied in the Tinderbox Hills, 25 km south of Hobart, Tasmania (Figure 2.1). The study site was a forest consisting of mature, almost pure *E. globulus* of an average height of 28 meters and juvenile seedlings of an average height of 65 cm with 4% of the juveniles beginning phase change from juvenile to mature foliage (Figure 2.2). Density of individuals did not vary significantly between cohorts, with the mature and juvenile cohorts exhibiting a density distribution of 0.012 and 0.010 individuals/$m^2$ respectively. This locality lies within the boundary of the Southeastern Tasmanian race of *E. globulus* as defined by Dutkowski and Potts (1999). The primary sample site consisted of a circle with a diameter of 140 m, within which all mature (168 individuals) and juvenile seedlings (115 individuals) were sampled. Two individuals within the sample site were close to death and DNA was not successfully extracted from these individuals. Sufficient amplification of DNA was also not successful in another three mature trees within the site, despite repeated DNA cleanup steps. Three *E. amygdalina* trees were also situated within the sample site. This species is from a different subgenus and not able to hybridize with *E. globulus* (Potts and Wiltshire 1997).

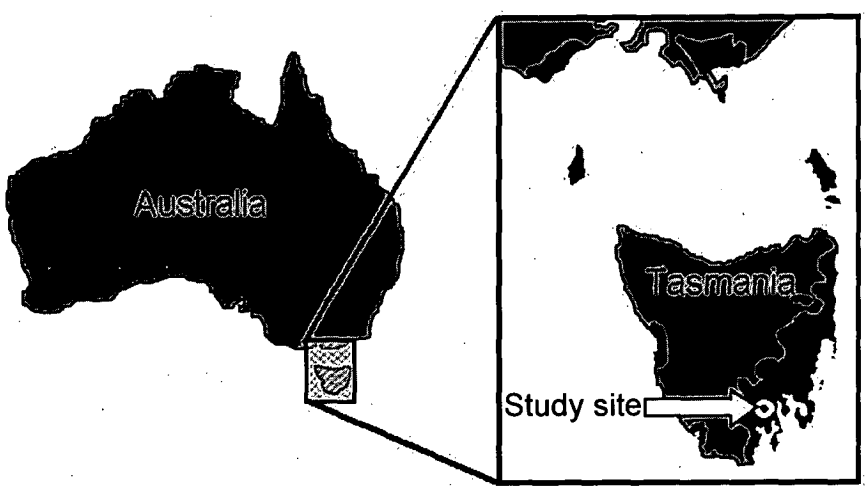



Figure 2.1: The position of the study site (lat. 43'2", long. 147'19", altitude ~300 m) within the native distribution (shaded black) of *E. globulus* is indicated. Image reproduced with permission from G. Dutkowski.

The mature trees and juvenile seedlings sampled from within the primary sample site are referred to as the 'mature cohort' and the 'juvenile cohort' respectively, in the following text. In addition, three transects of 10 mature *E. globulus* trees were sampled at 100 m, 200 m and 400 m from the center of the sample site in a northerly direction, with the transects running from east to west. These trees are referred to as the 'outliers' in the following text.

Leaf tissue of 106 mature trees and 115 juvenile seedlings was collected for DNA extraction, with a further 92 mature trees sampled by collecting cambial tissue. Initial sampling of 86 individuals was carried out in July 2001 with a further 112 mature individuals and 115 juvenile individuals completed in January 2003. Cambial tissue was collected by removing a 10 x 50 mm section of bark from the cortex and scraping the cambial layer with a clean razor blade, transferring the tissue into a microtitre tube containing 500 µl of CTAB buffer (Doyle and Doyle 1990) for transport and extraction.

Figure 2.2: The study site consisted of an open, almost pure stand of *E. globulus* on a gently sloping hill, with a westerly aspect. Average tree and seedling height was around 28 m and 65 cm respectively. The site is typical of Tasmanian dry sclerophyl forest (Duncan and Brown 1985), relatively exposed to the prevailing southwesterly winds and shows evidence of high fire frequency.

**Molecular Methods**

Total genomic DNA was extracted following the protocol of Doyle and Doyle (1990) using an incubation temperature of 55 °C. Whereas leaf tissue was ground before extraction as suggested by Doyle and Doyle (1990), cambial tissue was processed without grinding. The cambial scrape resulted in a mass of moist, disrupted cambial cells of around 40 µl in volume and DNA was extracted directly from the original sampling mix of cambial tissue and 500 µl CTAB buffer. DNA quantity and quality was assessed using either agarose gel electrophoresis with a standard molecular weight marker (lambda *Hin*dIII) or by specrophotometry.

Initially, 14 microsatellite loci where amplified for a subset 86 individuals. Nine highly polymorphic loci were selected (based on their PCR reliability and information content) for genotyping the complete population. Five of the nine microsatellite loci (EMCRC 2, EMCRC 7, EMCRC 8, EMCRC 10, EMCRC 11) used in this study were developed for *E. globulus* by Steane *et al.* (2001). The remaining four primer pairs (EMBRA 10, EMBRA 11, EMBRA 12, EMBRA 17) were developed by Brondani *et al.* (1998) and applied to *E. globulus* by Bundock *et al.* (2000). A combination of single primer pair PCR reactions and multiple primer pair (multiplex) PCR reactions were used to amplify the microsatellite loci, with the single locus PCR reaction conditions used for the "EMCRC" and "EMBRA" primers explained in Jones *et al* (2002) and Brondani *et al.* (1998) respectively. Fragment length of the EMCRC 8, EMBRA 11 and EMBRA 12 loci was determined using electrophoretic separation on a denaturing polyacrylamide gel, as described by Jones *et al.* (2002), using the Gel-Scan 2000 real-time acrylamide gel system (Corbett Research). Allele size was determined with both the Gene-Scan 350 (TAMRA) size standard and the Promega (CXR) 60-400 base fluorescent ladder, using the Gene Profiler $^{TM}$ software (Scanalytics, Inc.). Calibration between both size standards was necessary due to a difference in up to 3 base pairs between common fragment sizes of each standard. A varying number of individual samples were repeated from one gel run to the next (co loaded as an 'allelic ladder') to allow another standardization of allele size, independent of the commercial molecular size standard.

Two primer sets (1: EMCRC 2, EMCRC 7, EMCRC 10 and EMCRC 11, 2: EMBRA 10 and EMBRA 17) were developed for multiplex PCR. For both primer sets, the PCR reaction mix (final volume of 12.5 µl) contained two units of Taq Polymerase with 1x reaction buffer (67 mM Tris-HCl (pH 8.8), 16.6 mM $(NH_4)_2SO_4$, 45% Triton X-100, 0.2 mg/ml gelatin); 100 µg/ml BSA (bovine serum albumin); 120 µM of each dNTP, 1.5 mM MgCl; 1.3 M betaine; 0.0136% v/v DMSO; Primer set 1 (0.20 µM EMCRC 2 F and R, 0.40 µM EMCRC 11 F and R, 0.08 µM EMCRC 7 F and R, 0.08 µM EMCRC 10 F and R) or Primer set 2 (0.12 µM EMBRA 17 F and R , 0.20 EMBRA F and R); 50 ng genomic DNA. EMCRC 2, EMCRC 7, EMCRC 10, EMCRC 11, EMBRA 10 and EMBRA 17 were labeled with WellRED (Proligo, http://www.gensetoligos.com/) dyes D4-PA, D4-PA, D3-PA, D3-PA, D4-PA and D2-PA respectively. Multiplex PCR products were run on the Beckman Coulter™ CEQ™ 8000 Genetic Analyses System (loading 1 µl of product in 25 µl of deionised formamide) and sized against the CEQ™ DNA standard kit 400. The data was processed with the associated Beckman Coulter™ CEQ™ 8000 Genetic Analyses System software. If the multiplex reaction was not successful (around five percent of runs), loci for the sample were individually amplified and co loaded (coplexed) on the Beckman Coulter™ CEQ™ 8000 Genetic Analyses System.

**Data Analysis**

*Descriptive Analysis*

The 30 outlying individuals were only included in an initial spatial genetic autocorrelation to provide some insight into genetic differentiation at distances greater than that within the primary sample and were excluded from further analyses. For this reason, descriptive statistics are only displayed for the two cohorts within the primary sample site, so that they are representative of the individuals that all other analyses are based on.

Due to the difficulty in reliably testing significance of deviation from Hardy Weinberg equilibrium in allele frequencies, especially in highly polymorphic loci with multiple

rare alleles (e.g. Montoya-Delgado *et al.* 2001), deviation from Hardy Weinberg equilibrium was tested using both a $\chi^2$ goodness-of-fit test (Weir 1996) and Fisher's exact test (Weir 1996), in CERVUS (Marshall *et al.* 1998) and GDA 1.1 (Lewis and Zaykin 2001) respectively. A conservative approach was taken to assessing deviation in Hardy Weinberg equilibrium by removing loci only if they were found to deviate significantly in both cohorts with both tests (see Results). Potential linkage between loci was tested in both cohorts using Fishers exact test in GDA 1.1 with the significance level corrected for multiple comparisons (Rice 1989). The mean number of alleles per locus, allelic richness (i.e. the average allele number, standardized for variation in the number of individuals within a cohort, El Mousadik and Petit 1996), observed and expected heterozygosity, and Wrights fixation index ($F_{st}$) and inbreeding coefficient ($F_{is}$) (Weir and Cockerham 1984) between cohorts were calculated using FSTAT Version 2.9.3.2 (Goudet 1995).

*Spatial genetic correlations*

Initial correlation between spatial and genetic distance (Peakall *et al.* 1995) was tested in the mature cohort and outliers using the Mantel's test (Smouse *et al.* 1986) with the population genetic software GenAlEx V5 (Peakall and Smouse 2001). This analysis was repeated for the mature cohort alone, the juvenile cohort alone and the combined data set of both the mature and juvenile cohorts. Having found a significant global correlation in each sample, spatial genetic autocorrelation analysis was undertaken using GenAlEx V5 (Peakall and Smouse 2001). The genetic distance of measure of Peakall *et al.* (1995) was used, testing for spatial genetic autocorrelation at five meter distance intervals, using the maximum possible number of distances classes. This provided an illustration of the decay in genetic distance associated with increasing distance between individuals. As recommended by Peakall and Smouse (2001), 999 permutations were used in the calculation of standard error (95% confidence interval) for each analysis. The outliers did not significantly contribute any significant spatial genetic correlation and were not included in further analyses.

In an attempt to specifically quantify the spatial genetic autocorrelation in terms of relatedness and levels of bi-parental inbreeding, spatial genetic correlation of the mature, juvenile and combined cohorts was repeated using Lynch and Ritland's (1999) relatedness estimator (r) and Lynch and Ritland's (1999) fraternity, or 'four gene' relatedness estimator (Δ). This was done using the program SPAGEDI (Hardy and Vekemans 2002) which also allows spatial genetic analysis using an additional range of genetic distance measures: genetic distance (Rousset 2000); kinship (Loiselle *et al.* 1995; Ritland 1996); relatedness (Queller and Goodnight 1989; Hardy and Vekemans 1999; Wang 2002); fraternity (Wang 2002). The Lynch and Ritland (1999) estimators were selected for use primarily because the proportion of the Lynch and Ritland (1999) relatedness coefficient that is potentially due to bi-parental inbreeding and full-sibship can be dissected and expressed as the 'four gene' relatedness coefficient (Lynch and Ritland 1999).

The presence of multiple directions of spatial genetic variation was explored by firstly ordinating individuals based on their pair-wise genetic distance (Peakall *et al.* 1995; Smouse and Peakall 1999), using principal coordinate analysis (PCO, Sokal and Rohlf 1981) carried out with GenAlEx V5 (Peakall *et al.* 1995; Peakall and Smouse 2001). PCO allows the determination of the position of individuals along independent maximum variance axes that summarize multidimensional genetic variation within a population. Informative, independent PCO axes were then used in spatial autocorrelation analysis as described above to investigate spatial association of particular directions in genetic variation. Specifically, this was carried out by using the pair wise matrix of distance between individuals in PCO value along a particular axis in place of the standard genetic distance matrix in GenAlEx V5.

*Spatial patterns of genetic variation*

While a significant geographic association with genetic similarity had been identified by the spatial genetic autocorrelation analysis (see Results), there were no visible clues as to the manner in which this genetic variation was actually distributed or partitioned within the sample site. Consequently, no meaningful hypothesis regarding the potential

group structure could be tested using group-wise genetic distance measures. To overcome this problem, a Bayesian, model based approach was used in an attempt to define genetically distinct groups without a prior hypothesis. The software STRUCTURE (Pritchard *et al.* 2000a) was used to analyze the mature, juvenile and combined cohorts. This program assigns individuals to groups based on their genotypic similarity, while simultaneously calculating gene frequencies of those groups and probabilistically reassigning individuals to the newly refined gene frequency based groups. STRUCTURE repeats this process many times using a Markov Chain Monte Carlo method (MCMC) until stabilization indicates the model of best fit. Before the MCMC simulations are run, it is necessary to estimate the number of groups within the data. The appropriate number of groups in each cohort was investigated by calculating the posterior probabilities for one to 10 groups, using a burnin period of $1x10^5$ and $1x10^5$ MCMC repetitions, assuming correlated gene frequencies, admixture and providing no prior information, repeating each run three times, analyzing each cohort separately. The mean log likelihood value over the three repeats was plotted for each group number, with the highest log likelihood indicating the optimal number of groups (Pritchard *et al.* 2000a). The STRUCTURE analysis was then repeated with a burnin period of $1x10^5$ and $1x10^6$ MCMC repetitions, clustering individuals to the appropriate number of groups in each cohort. The spatial distribution of the resulting groups was illustrated by labeling the geographic position of all individuals with its assigned group.

The percentage of genotypic variation accounted for by the STRUCTURE groups in both cohorts was determined in an analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992) using GenAlEx V5 (Smouse and Peakall 1999). The genetic similarity between individuals in each group was also investigated by plotting their position along the two most informative principal coordinate axes from the initial principal coordinate analysis and labeling the individuals with the appropriate STRUCTURE assigned group.

The spatial distribution of genetic variation along each of the PCO axes that displayed significant spatial autocorrelation (transcending the confidence limits of the spatial autocorrelogram) in each cohort was visualized by using a three-dimensional contour plot of the most informative principal component (z axis), against geographical position (x and y axis), using an inverse distance smoothing factor (sampling proportion = 0.1,

exponent = 1, SigmaPlot 2000 for Windows version 6). This approach was similar to that used by Hanotte *et al.* (2002) for example. Due to the sample size constraints of GenAlEx V5, Genstat 5 Release 3.2 (1995) was used to carry out a principal coordinate analysis of the combined juvenile and mature cohort dataset. The combined analysis of the two cohorts allowed the spatial pattern of variation in gene frequencies to be compared across cohorts, using the same PCO axis to differentiate individuals. A more extreme smoothing factor was applied to this PCO (inverse distance, sampling proportion = 0.3, exponent = 1, SigmaPlot 2000 for Windows version 6), allowing comparison of the spatial position of contours (isopleths) between the cohorts. This provides an easily interpreted visualization of the geographic movement in common gene frequency combinations between cohorts, and so illustrates realized, directional gene flow in this population.

To quantify the genetic variation geographically bounded by the PCO isopleths, pair-wise matrices of Lynch and Ritlands (1999) mean pair-wise relatedness, using SPAGEDI (Hardy and Vekemans 2002) and the coancestry coefficient $\theta$, using GDA 1.1 (Lewis and Zaykin 2001), otherwise known as the pair-wise $F_{st}$ (Weir and Hill 2002) were calculated in each cohort. The coancestry coefficient $\theta$ was calculated because the software provides a significance of group-wise comparisons associated with this coefficient.

## Results

### Genetic diversity

All nine microsatellite loci were highly polymorphic with the number of alleles per locus ranging from 12 to 21, averaging 17.6 (Table 2.1). One locus, EMCRC 10 (not shown), showed significant deviation from Hardy Weinberg equilibrium in both cohorts, most likely due to a high frequency of null alleles and was not used in the analysis. The presence of null alleles was not detected in the other loci. The remaining eight loci were suitable for further analysis and the number of alleles, allele size range, expected and observed heterozygosity for each locus of these loci are shown in Table 2.1. No significant linkage between the eight remaining loci was detected in this population.

Very little difference between cohorts was seen in expected and observed heterozygosity (Table 2.2). Observed heterozygosity values for the mature cohort, juvenile cohort and overall population were 0.85, 0.83 and 0.84 respectively. These values did not differ markedly from the expected heterozygosity, with $F_{is}$ close to zero in each case. The juvenile cohort displayed slightly lower allelic diversity, as indicated by a lower average number of alleles and lower allelic richness (Table 2.2). Like-wise, the juvenile cohort displayed a lower number of private alleles per locus. This may suggest that either, the progeny are derived from a reduced number of mature individuals, the progeny are slightly inbred, or that out-crossing was more prevalent when the mature cohort was initially established. There was however, no significant overall genetic difference between the two cohorts as indicated by an $F_{st}$ value of $0.000 \pm 0.001$ (jackknifing over loci). While slight variation is suggested by allelic richness and the numbers of private allele in each cohort, on a more general scale, there has been no significant inbreeding, or introgression of other alleles into the juvenile cohort. Only marginal difference in $F_{is}$ between cohorts was evident.

Table 2.1: All loci used in the analysis were highly informative. Locus name, number of alleles (N° alleles), allele size, expected heterozygosity ($H_e$), observed heterozygosity ($H_o$) for each locus are shown.

| Locus | N° alleles | Allele size | $H_e$ | $H_o$ |
|---|---|---|---|---|
| EMCRC 11 | 16 | 223-255 | 0.84 | 0.82 |
| EMCRC 7 | 20 | 175-317 | 0.80 | 0.75 |
| EMCRC 2 | 12 | 155-189 | 0.78 | 0.78 |
| EMCRC 8 | 16 | 231-263 | 0.87 | 0.87 |
| EMBRA 11 | 21 | 96-144 | 0.86 | 0.87 |
| EMBRA 10 | 20 | 114-152 | 0.88 | 0.88 |
| EMBRA 17 | 20 | 120-161 | 0.88 | 0.91 |
| EMBRA 12 | 16 | 108-174 | 0.86 | 0.89 |

Table 2.2: Genetic diversity within the mature and juvenile cohorts of the study population: $N$ = number of individuals; $\mu$ = mean number of alleles per locus; $R_t$ = allelic richness; $Pa$ = mean number of private alleles per locus; $H_e$ = expected heterozygosity; $H_o$ = observed heterozygosity; $F_{is}$ = Wright's inbreeding coefficient. Population = both cohorts combined.

| Cohort | $N$ | $\mu$ | $R_t$ | $Pa$ | $H_e$ | $H_o$ | $F_{is}$ |
|---|---|---|---|---|---|---|---|
| Mature | 168 | 15.88 | 14.52 | 2.88 | 0.84 | 0.85 | -0.011 |
| Juvenile | 115 | 13.75 | 13.72 | 1.63 | 0.85 | 0.83 | 0.018 |
| Population | 283 | 14.61 | 14.44 | *na* | 0.85 | 0.84 | 0.001 |

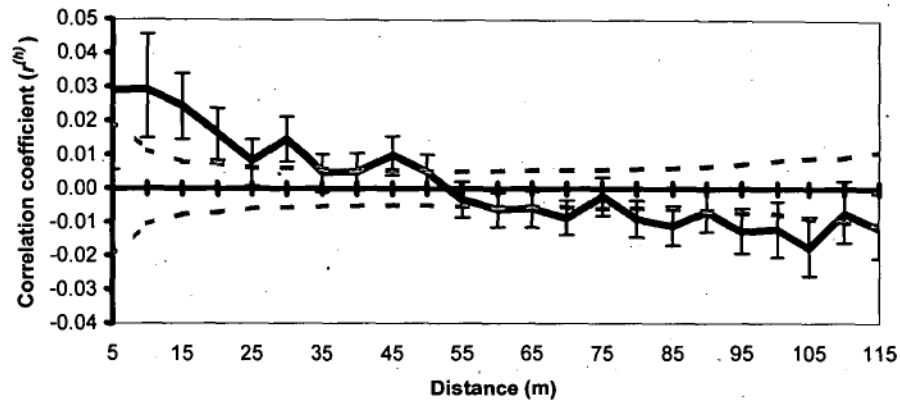
**Spatial genetic autocorrelation**

Within the combined mature cohort and outliers, a significant global correlation ($r$ = 0.073, P = 0.031) between genetic similarity (Peakall *et al.* 1995) and geographic distance was detected using the Mantels test. This was most likely caused by a significantly positive spatial autocorrelation coefficient from zero to 45 m (data not shown). Very little difference was seen in the autocorrelogram of the mature cohort with the outliers excluded, however the global correlation between genetic and geographic distance matrices was higher ($r$ = 0.133, P = 0.001). In this case, significantly positive spatial autocorrelation was displayed between a distance of zero and 35 m, with another significant correlation at a distance of 45 m (Figure 2.3a). While a significant correlation between geographic and genetic distance matrices was also evident in the juvenile cohort ($r$ = 0.099, P = 0.004), spatial genetic structure was weaker in this cohort (Figure 2.3b). A significantly positive autocorrelation was only present within the 10-20 m and 30-35 m distance classes. When analyzed as a combined

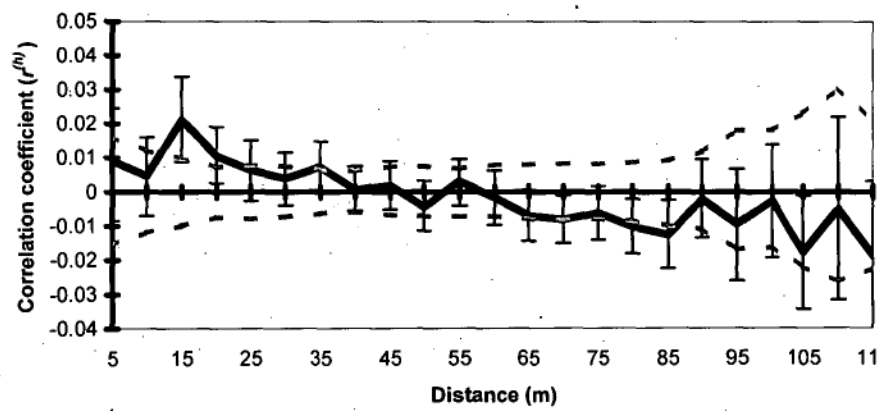
dataset, the two cohorts displayed an autocorrelation coefficient of intermediate value between that of the mature and juvenile cohorts alone ($r = 0.110$, $P = 0.001$, Figure 2.3c). More importantly, the slope of the correlation coefficient was very constant from 10 m to 100 m, suggesting a constant, clinal pattern of decline in genetic distance associated with increasing geographic distance.

In this case, un-related individuals (in the sense of identity by recent descent) should display a relatedness value of 0, or a negative value (the reason for which is explained in Lynch and Ritland 1999). Hence, the spatial genetic plots based on Lynch and Ritland's (1999) relatedness estimator, suggest a significantly higher relatedness between trees separated by up to 45 m in the mature cohort, between 10 m and 15 m in the juvenile cohort and up to 45 m in combined cohorts (Figure 2.4a,b,c). Some debate surrounds the interpretation of spatial autocorrelation and other spatial genetic plots, as to whether family patch size should be elucidated from the point where the correlation coefficient reaches 0, where the correlation coefficient reaches the inner bounds of the confidence limits, or where the gradient of the correlation coefficient becomes particularly steep. As Vekemans and Hardy (2004) indicate, the point at which the plot crosses the x axis, and that where it reaches the confidence limits around the mean, are dependent on the sample scheme of the study (Fenster *et al.* 2003). Accordingly, in this study, family patch size is only suggested when the plot decreases steadily until a point at which no further trend is detectable (as outlined in Vekemans and Hardy 2004). While the correlation is not particularly strong within the juvenile cohort (Figure 2.4b), a potential weak family grouping between 10 and 15 m in radius may exist. The mature cohort (Figure 2.4a) is a little more difficult to interpret, however a family group of around 55 m in radius may be evident as, at this point, the plot has just undergone a rapid decline in the spatial genetic association (Figure 2.4a), followed by a general plateau. Only a very slight change in slope of the plot of relatedness with distance around 55 m was evident in the combined cohort analysis, suggesting no clear family group structuring, and instead, a reasonably steady decline in relatedness with distance (Figure 2.4c).

(a) Mature cohort




(b) Juvenile cohort




(c) Mature and juvenile cohort

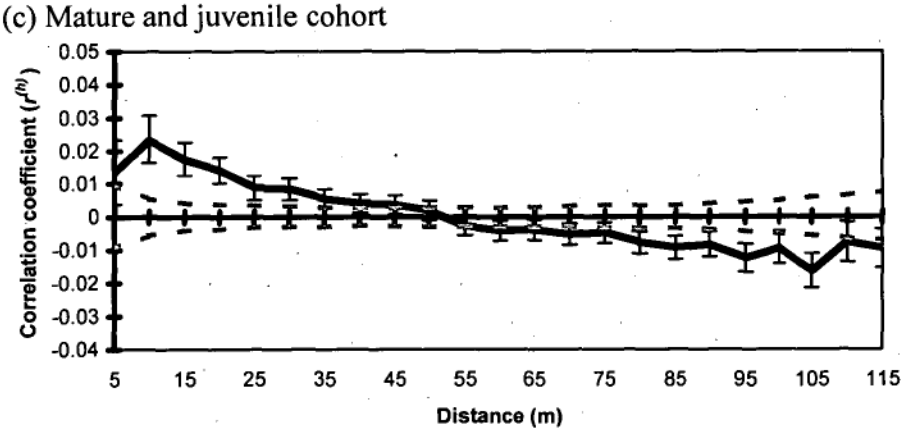
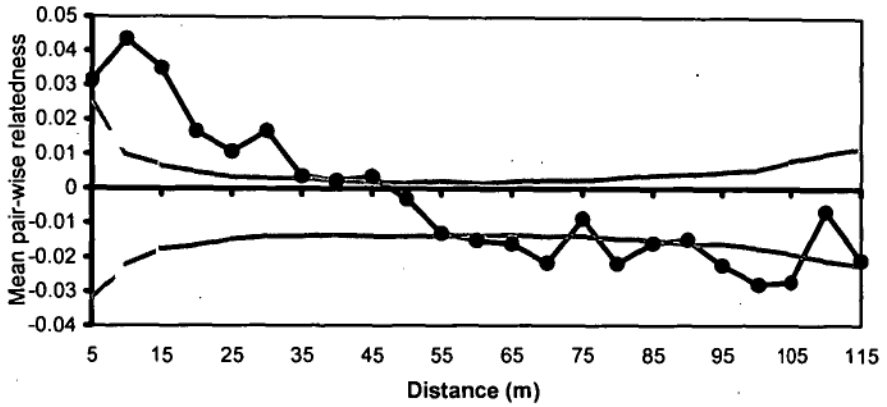
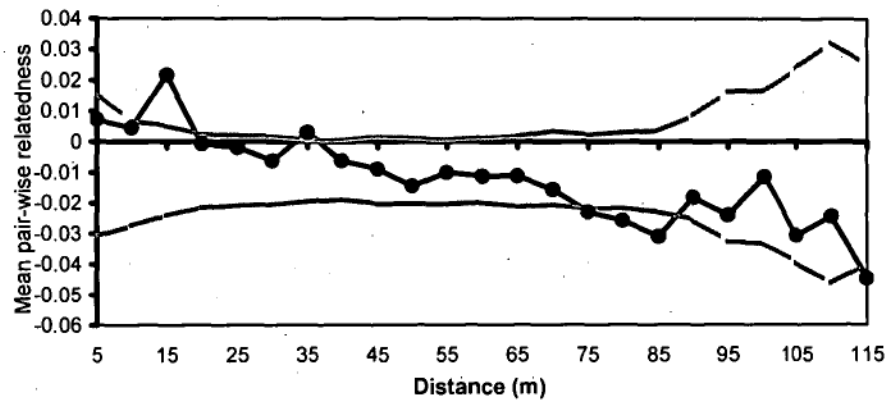



Figure 2.3: Multilocus spatial-genetic autocorrelogram of genetic distance (Peakall *et al.* 1995) for trees within the primary study site. Spatial genetic structure is shown in the mature cohort (**a**), juvenile cohort (**b**) and both cohorts combined (**c**). 95% null hypothesis confidence regions around zero are marked by grey dotted lines. Standard error bars for each distance classes are shown.

(a) Mature cohort




(b) Juvenile cohort




(c) Mature and juvenile cohort

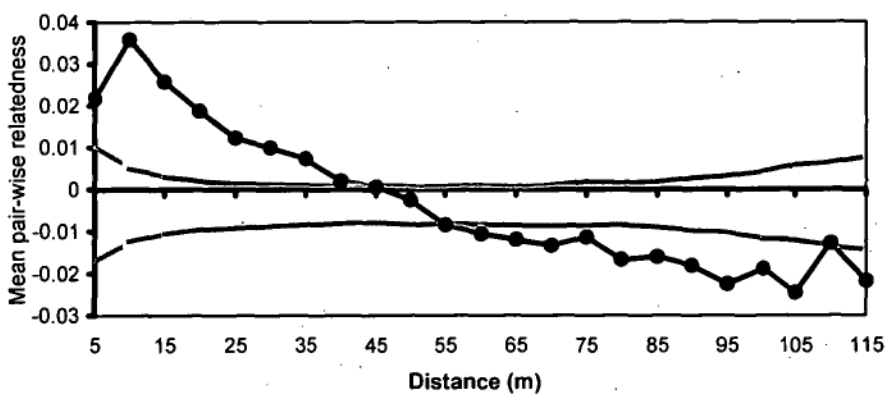
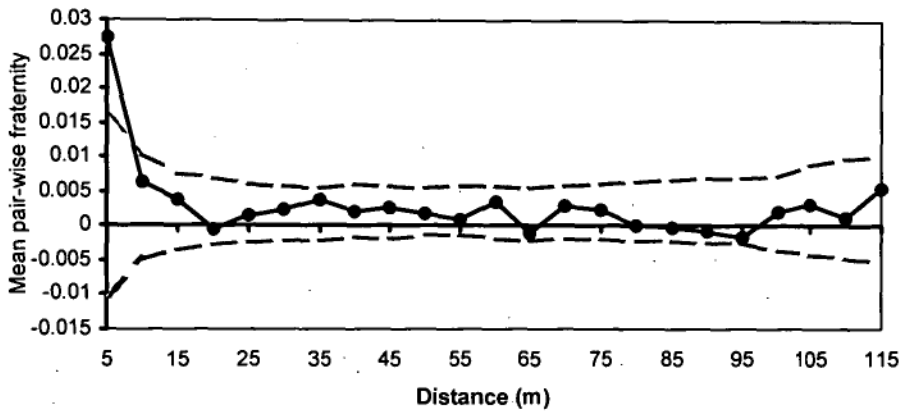
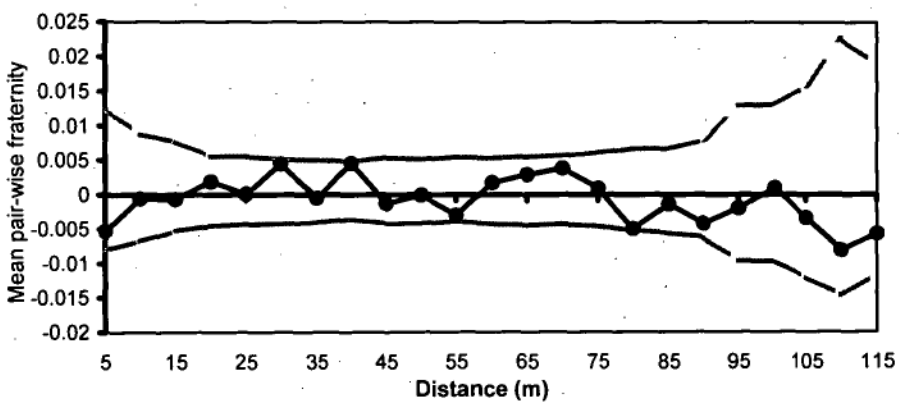



Figure 2.4: Spatial-genetic association based on Lynch and Ritland's (1999) relatedness estimator. A steep decline in the relatedness with distance is seen in the mature cohort (**a**). The juvenile cohort (**b**) displays a much weaker association, with a mean pair-wise relatedness value of 0 displayed by trees separated by a distance above 20 m (excluding the 30-35 m distance class that registers a marginally significant positive association). The combined analysis of both cohorts (**c**) displays an almost linear clinal pattern of decrease in relatedness with distance. 95% null hypothesis confidence regions around the mean are marked by gray dotted lines.

(a) Mature cohort




(b) Juvenile cohort




(c) Mature and juvenile cohort

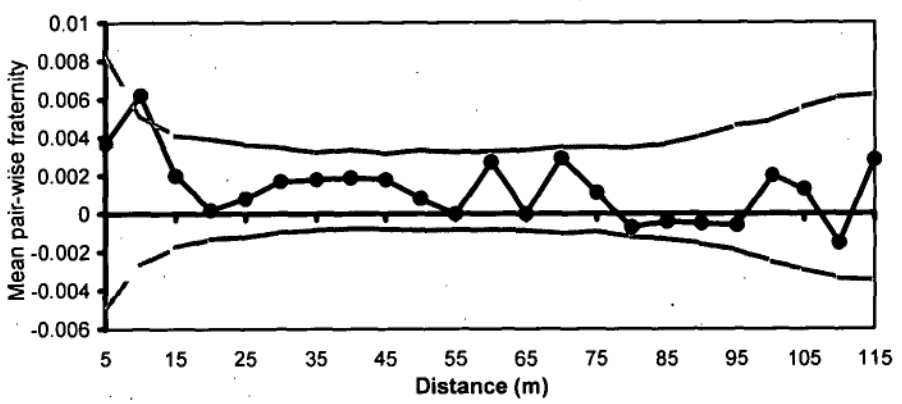



Figure 2.5: Spatial-genetic association based on Lynch and Ritland's (1999) fraternity coefficient (4 gene relatedness estimator). Bi-parental inbreeding and/or full-sibship was significantly higher between trees within 0 – 5m of each other than those separated by greater than 5 m in the mature cohort (**a**). No such association was seen in the juvenile cohort (**b**) with the mean pair-wise fraternity coefficient for each distance class never transcending the 95% confidence intervals (indicated by gray dotted lines). A slightly positive association is displayed in the combined analysis of both cohorts (**c**), however, the mean fraternity value that is seen for the 10 –15 m distance class is very low (0.006).

A more specific comparison of the association between relatedness and geographic distance in each cohort can be made by comparing the slope of the regression of each plot (Vekemans and Hardy 2004). In this case, the regression of relatedness against the logarithm of geographic distance of the mature cohort, over distances under 55 m (the distance beyond which no clear trend is present in the mature and juvenile cohorts), resulted in a regression coefficient of -0.023 which was more than twice that of the juvenile cohort (-0.010). As the plots suggest, the combined analysis displayed an intermediate slope of -0.018.

It should be noted that within the mature cohort, pairs of trees of between five and 10m apart displayed the highest average relatedness value of 0.044 (Figure 2.4a). This value is below that expected of 4$^{th}$ degree relationships (e.g. 2$^{nd}$ cousins, r = 0.061), however this average relatedness value was derived from 213 pair-wise comparisons, of which almost 10% displayed a relatedness value of above 0.25 (indicative of at least grandparent parental, half sibling or avuncular relationships). In comparison, only four percent of all pair-wise relationships over the entire mature cohort displayed this relatedness value or above. The maximum pair-wise relatedness value within the mature cohort was 0.86, found between two trees that were only separated by a distance of several metres.

The component of relatedness that is due to either bi-parental inbreeding or full-sibling relationships can be estimated by Lynch and Ritland's (1999) fraternity measure. Within the mature cohort (Figure 2.5a), it is clear that a significant proportion of the high relatedness values seen between trees zero to five meters apart may be due to bi-parental inbreeding and full-sibling relationships. However, following a rapid decline in mean pair-wise fraternity, there is virtually no detectable evidence of this type of bi-parental inbreeding and full sibling relationship beyond five meters using this approach. Due to very high values of pair-wise relatedness (0.66 and 0.60) and very low levels of pair-wise fraternity (-0.14), it is quite likely that at least two individuals that are within zero to five meters of each other share parent offspring relationships.

No such evidence for a spatial association of bi-parental inbreeding or full sib-ship between individuals within the juvenile cohort was detected using this method, as mean

pair-wise fraternity is not higher between close trees than distance ones (Figure 2.5b). When the mature and juvenile cohorts were analyzed as a single group, some spatial correlation with fraternity may be indicated by a marginally significantly higher mean pair-wise fraternity between trees separated by five to 10 m (Figure 2.5c), however the low mean pair-wise fraternity value of 0.006 between trees separated at this distance suggests that this is not a strong association.

**Spatial patterns of genetic variation**

Using the Bayesian based STRUCTURE software, the distribution of likelihood values in the mature cohort did not display a distinctly higher posterior probability for any one particular number of genetic groups, with between three and seven groups suggested by sharing of very similar, relatively high, likelihoods (Figure 2.6a). When individuals were assigned to between three and seven groups by STRUCTURE, they were unevenly allocated in number to the modeled groups. In addition, there was high consistency between likelihood values for repeated runs with the same number of groups (Figure 2.6a). Both of these phenomena are clear prerequisite indicators of significant genetic structuring within the mature cohort (Pritchard *et al.* 2000a). Evanno *et al.* (2005) suggest that the second order rate of change of the likelihood function with respect to the particular number of groups ($\Delta K$) may indicate the real number of groups in their simulated datasets. When this was carried out, a peak at both the three and six group solution occurred, providing little additional information.

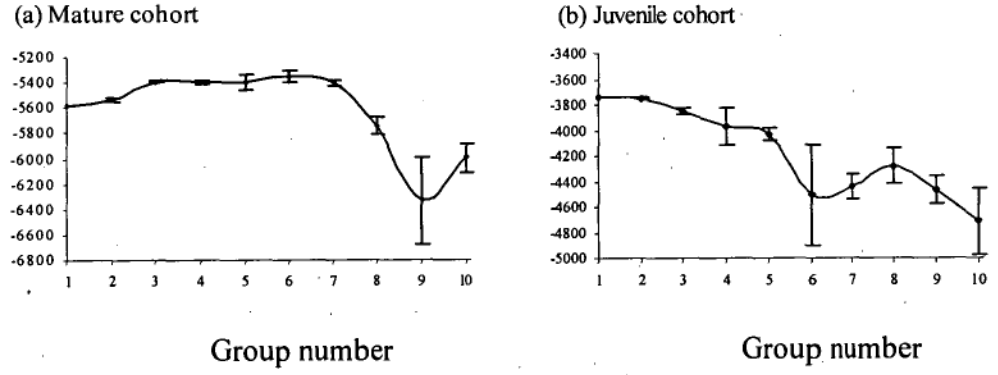



Figure 2.6: The distribution of mean log likelihood values of the mature (**a**) and juvenile (**b**) cohorts for a number of groups between 1 and 10, with standard errors included. Likelihood values were averaged over 10 runs. Poor differentiation between estimates of the optimal number of groups within the range of 3 to 7 is indicated in the mature cohort, while the likelihood of a particular number of groups above 2 rapidly decreases in the juvenile cohort.

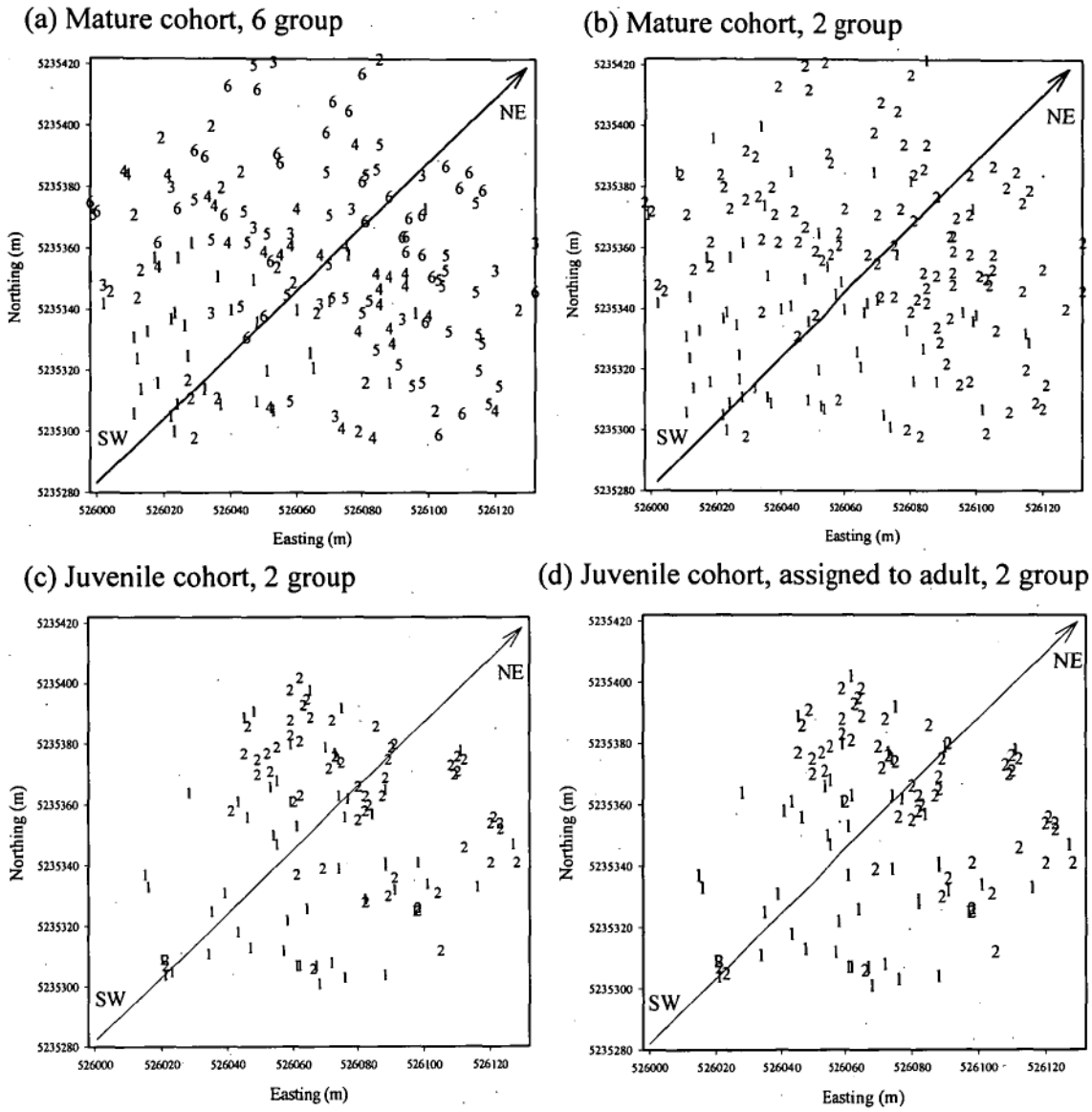



Figure 2.7: Geographic position of individuals, labeled with their group defined by the STRUCTURE analysis: **a)** mature cohort clustered into 6 groups which had the highest posterior probability in the initial STRUCTURE analysis; **b)** mature cohort clustered into two groups to summarise the geographic direction of genetic variation in the cohort; **c)** juvenile cohort clustered into two groups, which (aside from one-group) had the highest posterior probability in the initial STRUCTURE analysis, displaying a similar pattern to the two-group mature analysis; **d)** juvenile allocated to the two groups defined by the two-group mature analysis.

Whilst not significantly greater, the six-group solution did show the highest mean likelihood and was subsequently chosen as the optimal group number for further analysis, in compliance with the guidelines suggested by Pritchard *et al.* (2000a).

When an ANOVA of the six-group solution was carried out, the six groups accounted for 12.3% ($P < 0.001$) of the overall variation within the mature cohort. When the six

STRUCTURE groups are spatially displayed, a clear geographic pattern to the grouping is evident, with the groups (labeled 1 to 6) arranged in a southwesterly direction (Figure 2.7a). Although, when following the guidelines of Pritchard *et al* (2000), the two-group STRUCTURE solution does not optimally describe the genetic structure within the mature cohort, it does effectively summarizes the spatial genetic differentiation along the same southwesterly direction (Figure 2.7b). In this case, individuals from group 1 tend to cluster in the southwesterly corner of the population sample, with those from group 2 situated towards the northeast of the sample site.

The STRUCTURE analysis of the juvenile cohort produced the highest likelihood values for the one and two group solutions (Figure 2.6b). The ANOVA of the juvenile cohort indicated that variation between the two STRUCTURE groups accounted for 6.1% ($P < 0.001$) of the total genotypic variation within the cohort. When the two-group STRUCTURE solution is spatially mapped (Figure 2.7c), a clear non-random spatial distribution along the southwest vector is displayed, in a similar manner to that of the mature cohort. When individuals from the juvenile cohort were assigned to the two groups predefined in the mature cohort by STRUCTURE, a similar geographic pattern resulted (Figure 2.7d) indicating a close similarity between the variation in gene frequency combinations that define the two groups in both cohorts.

Principal coordinate analysis (PCO) clearly illustrates the genetic relationship between the groups defined by STRUCTURE in both the mature and juvenile cohort. The first two principal coordinates explain 7.6% of the genetic variation in the mature cohort (Figure 2.8a). In this space, there is continuous variation at the individual level and the structure groups 1 and 6 are the most genetically extreme, with the other groups being intermediate within these. While the principal coordinate axes plotted in Figure 2.8a indicate distinct genetic variation amongst the groups defined by STRUCTURE, they only explain a small proportion of the total genetic variation within the cohort, and so suggest a subtle level of directional, continuous genetic differentiation, within a highly stochastically variable population. However, within the mature cohort, further investigation of the relationship between the six STRUCTURE groups and the first 20 principal coordinate axes, using a one-way ANOVA (PROC GLM, SAS V8) to test each

axis, revealed a number of additional associations (Table 2.3). While PCO axis 1 clearly suggests that the six-group STRUCTURE model effectively dissects a potential genetic

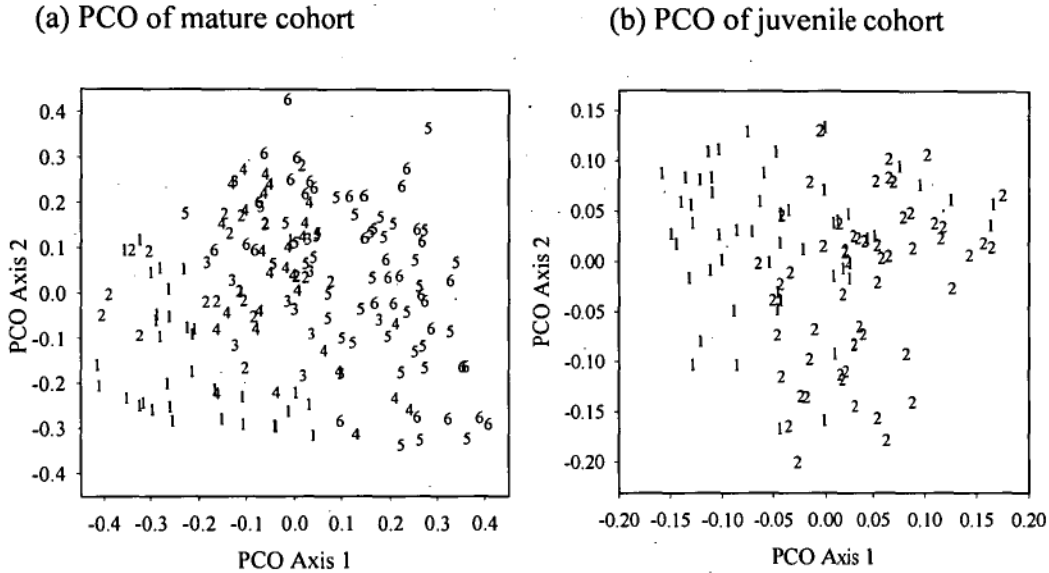



Figure 2.8: Separate principal coordinate analysis of mature (a) and juvenile (b) cohorts, based on the genetic distance measure described in Smouse and Peakall (1999), show continuous genetic variation in each cohort and the relationship between each STRUCTURE defined group. Individuals are labeled with the STRUCTURE groups as defined by the 6 group solution for the mature cohort and the 2 group solution for the juvenile cohort.

continuum along this axis, PCO axis 2 indicates that group 1 significantly differs from all other groups in another, independent way (Table 2.3). This can only be caused by a gene frequency combination that is specifically unique to that group. Likewise, unique gene frequency combinations that significantly characterize group 4 and group 2 are suggested by PCO axis 4 and PCO axis 6 respectively (Table 2.3). Further significant partitioning of the STRUCTURE groups by PCO axes 3, 5, and 7 indicate another three subtle directions in relationship between groups. Within the juvenile cohort, the variation is also continuous and the main difference between the two STRUCTURE groups occurs along the first principal coordinate axis, which accounts for 3.84% of the total variation between individuals (Figure 2.8b). In combination, the first two principal coordinate axes account for 7.3% of the genetic variation in this cohort. After the Bonferroni adjustment for multiple comparisons (Rice 1989), the two STRUCTURE

Table 2.3: The first 7 principal coordinate axes (PCO) indicated significant relationships between the 6 STRUCTURE groups (after the Bonferroni adjustment for multiple comparisons, Rice 1989) in the mature cohort. Eigen values, F values and the probability (Prob) of the association between the PCO axis and the six-group STRUCTURE solution are shown. Significantly different STRUCTURE groups (based on each independent PCO axis) do not share a similar letter (Tukey–Kramer test, PROC GLM, SAS V8).

| | | | | STRUCTURE GROUPS | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PCO axis | Eigen value | F value | Prob | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 0.956 | 47.26 | <0.001 | c | c | b | b | a | a |
| 2 | 0.766 | 7.44 | <0.001 | b | a | a | a | a | a |
| 3 | 0.720 | 6.41 | <0.001 | bc | ab | abc | cb | c | a |
| 4 | 0.644 | 24.78 | <0.001 | a | a | a | b | a | a |
| 5 | 0.583 | 13.04 | <0.001 | ab | bc | bc | ab | a | c |
| 6 | 0.545 | 7.3 | <0.001 | a | b | a | a | a | a |
| 7 | 0.491 | 7.67 | <0.001 | bc | a | c | cb | ab | c |

groups could be independently differentiated along PCO axes 1, 2 and 5 (data not shown). This indicates that at least three independent gene frequency combinations characterize the variation between the two groups. The combined analysis of both cohorts displayed a similar pattern of continuous, directional genetic variation throughout the sample area, in a very similar fashion to the mature cohort, but with slightly less extreme variation (data not shown). No comparisons between PCO axes in this case are included because no further information in addition to that of the single cohort studies was gained from this analysis.

To further explore the spatial distribution of the genetic variation expressed by the principal coordinate analysis, spatial autocorrelation of each of the first 10 principal coordinate axes within each cohort was undertaken (the principal coordinates beyond the first 10 accounted for insufficient variation to warrant further investigation). PCO axes that displayed a significant association with distance were then spatially interpolated, with the PCO axis plotted as the z axis against the geographic position of the individual trees or juveniles (x and y axis) (Figures 2.9 and 2.10). Four groups of genetically distinct individuals were revealed in the mature cohort by a significant spatial autocorrelation in the first, second and fourth PCO axes (Figure 2.9). Interpolation of the first PCO axis in the mature cohort suggests that the genetic continuum initially indicated potentially consists of two major groups (blue-green and red-orange, Figure 2.9a), characterized by plateaux in the interpolated PCO contours, divided by a region of steeper contours indicative of rapid genetic change.

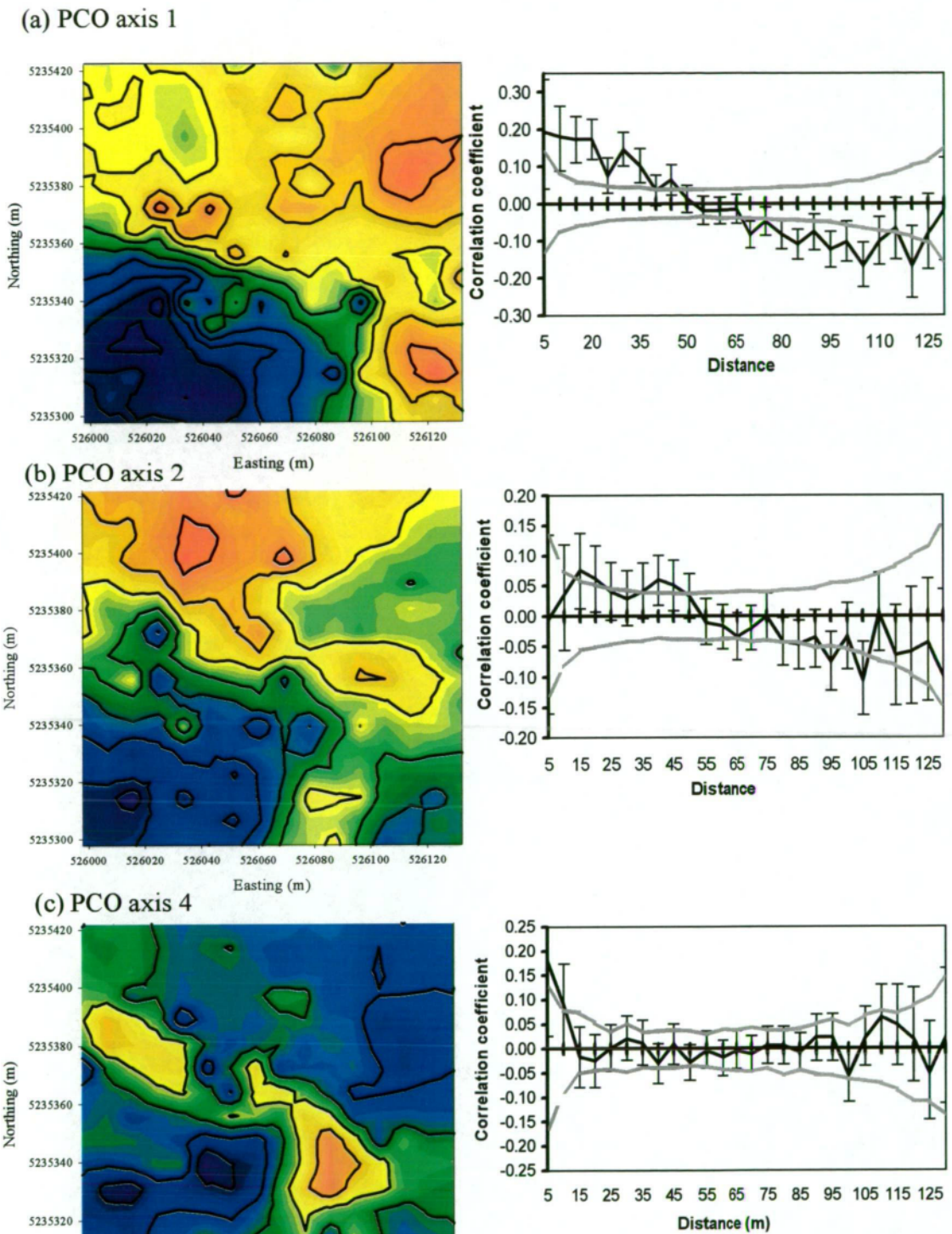



Figure 2.9: Spatially interpolated contour maps (left) of the PCO axes that displayed significant spatial autocorrelation (right) in the mature cohort. While PCO axes 1 (**a**) and 2 (**b**) display 3 widely dispersed groups (indicated by plateaux in PCO contours, and consistent with the 3 group structure analysis not shown), PCO axis 4 (**c**) displays two narrow patches of genetically similar (orange/red) individuals. (Colour coding shown in Figure 2.10).

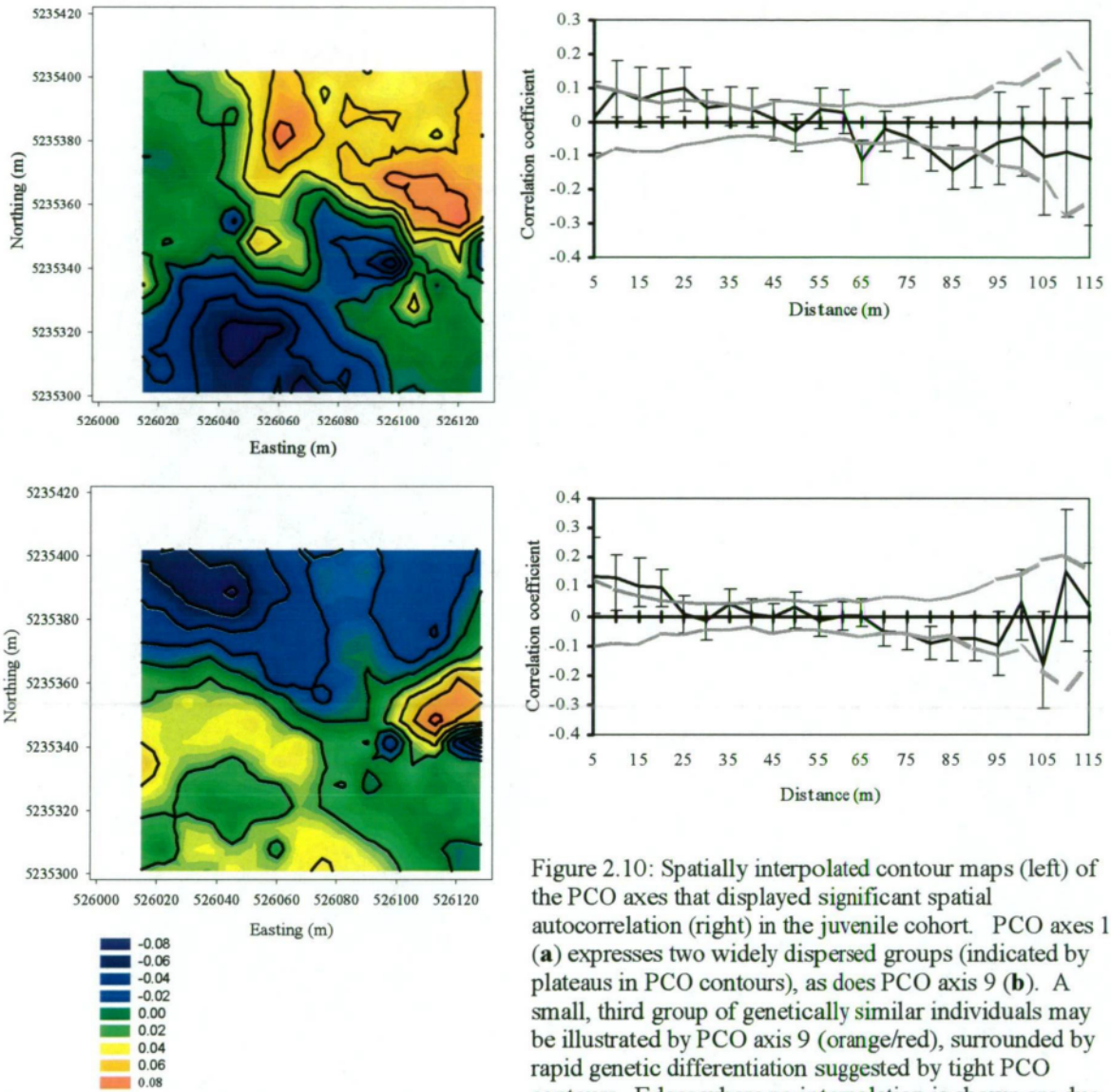



Figure 2.10: Spatially interpolated contour maps (left) of the PCO axes that displayed significant spatial autocorrelation (right) in the juvenile cohort. PCO axes 1 (**a**) expresses two widely dispersed groups (indicated by plateaus in PCO contours), as does PCO axis 9 (**b**). A small, third group of genetically similar individuals may be illustrated by PCO axis 9 (orange/red), surrounded by rapid genetic differentiation suggested by tight PCO contours. Edges where no interpolation is shown are due to lack of seedlings in these areas to provide sufficient data.

Comparison between the first and second PCO axes indicates that the northeastern group may actually consist of two genetically distinct groups, one tending toward the north and the other in the southeast, while the initial southwestern group remains coherent (Figure 2.9b). The fourth PCO coordinate displayed another genetically distinct group, running from the northwest to the southeast (Figure 2.9c). When comparing PCO axes 1 and 2,

this zone could be considered a zone of admixture between the groups defined by these axes (Figure 2.9), however, PCO axis 4 indicates that this group of individuals is genetically distinct in its own right.

Within the juvenile cohort, spatial interpolation of the two PCO axes (1 and 9) that displayed a significant spatial autocorrelation (Figure 2.10) revealed patterns that generally mirrored those already displayed by PCO axes 1 and 2 in the mature cohort (Figure 2.9 and 2.10). While the juvenile group has only been split into two groups by the STRUCTURE analysis, these two groups are still well visually defined by both PCO axes 1 and 9. However as indicated above, PCO axis 9 did not display a significant difference between the two structure groups.

When individuals from the mature cohort were grouped, based on the spatially interpolated PCO 1 contours (isopleths) that bound them (spanning a difference of 0.04 between each interpolated PCO contour) (Figure 2.9), the observed mean pair-wise relatedness values (Lynch and Ritland 1999) for each group (Table 2.4a) mirrored that predicted by the initial spatially associated decay in relatedness suggested by the autocorrelation analysis of the entire mature cohort (Figure 2.4a). The mean pair-wise relatedness expressed within the most southwesterly group (bound by the most extreme spatially interpolated PCO axis value of -0.08) was 0.106, which is just below that expected of cousins (r = 0.125, Lynch and Ritland 1999). This can be regarded as very high, considering it is a mean value of pair-wise relatedness between 22 individuals (20% of the entire sample). When the mean pair-wise comparisons between individuals in this group and the other spatially bound groups are calculated (Table 2.4a), the resulting group-wise relatedness values fall distinctly in relation to distance from the initial group. Coancestry coefficients for each group-wise comparison (a pair-wise $F_{st}$ value *between groups*, as opposed to between one group and the rest of the population, displayed a similar trend (Table 2.4a,b,c). It is interesting to note that while some groups expressed significant genetic differentiation to others, based on the coancestry coefficient (Weir and Hill 2002), they did not necessarily display high levels of *within group* relatedness (Lynch and Ritland 1999).

Table 2.4: Summary of the mean pair-wise relatedness values (Lynch and Ritland 1999) and coancestry coefficient (Weir and Hill 2002) (similar to a between group pair-wise measure of $F_{st}$) between each group bound by the contours of the significant spatially interpolated PCO axes (isopleths). A difference of 0.04 between contours defined each isopleth. Within group mean pair-wise relatedness (Lynch and Ritland 1999) is shown in the diagonal (italics). Between group mean pair-wise relatedness is shown below the diagonal and between group coancestry coefficient shown above the diagonal for the mature (**a,b,c**) and juvenile (**d,e**) cohorts. The significantly positive coancestry coefficient values ($P<0.05$, after adjustment for multiple comparisons) are indicated (*).

(a) Mature cohort PCO axis 1 (from Figure 2.9a)

| PCO contour boundary | < - 0.08 | -0.08 to -0.04 | -0.04 to 0.00 | 0.00 to 0.04 | > 0.04 |
|---|---|---|---|---|---|
| **< -0.08** | *0.106* | -0.002 | 0.022 | 0.050* | 0.095* |
| **-0.08 to –0.04** | 0.063 | *0.014* | 0.010 | 0.018 | 0.052* |
| **-0.04 to 0.00** | 0.026 | -0.002 | *0.008* | 0.008 | 0.038* |
| **0.00 to 0.04** | -0.0278 | -0.017 | -0.012 | *-0.004* | 0.014 |
| **> 0.04** | -0.053 | -0.030 | -0.023 | 0.003 | *0.033* |

(b) Mature cohort PCO axis 2 (from Figure 2.9b)

| PCO contour boundary | < - 0.04 | - 0.04 to 0.00 | 0.00 to 0.04 | > 0.04 |
|---|---|---|---|---|
| **< -0.04** | *0.097* | 0.013 | 0.047* | 0.077* |
| **-0.04 to 0.00** | 0.030 | *0.006* | 0.009* | 0.032* |
| **0.00 to 0.04** | -0.028 | -0.012 | *-0.003* | 0.006 |
| **> 0.04** | -0.055 | -0.027 | 0.002 | *0.038* |

(c) Mature cohort PCO axis 4 (from Figure 2.9c)

| PCO contour boundary | > 0.02 | - 0.02 to 0.02 | < -0.02 |
|---|---|---|---|
| **> 0.02** | *0.016* | 0.014* | 0.040* |
| **-0.02 to 0.02** | -0.011 | *-0.005* | 0.009* |
| **< -0.02** | -0.030 | -0.002 | *0.033* |

(d) Juvenile cohort PCO axis 1 (from Figure 2.10a)

| PCO contour boundary | < - 0.06 | - 0.06 to – 0.02 | - 0.2 to 0.02 | >0.02 |
|---|---|---|---|---|
| **< -0.06** | *0.046* | -0.003 | 0.006 | 0.051* |
| **-0.06 to –0.04** | 0.011 | *-0.016* | -0.005 | 0.019* |
| **-0.04 to –0.02** | 0.01 | -0.008 | *-0.009* | 0.018* |
| **> 0.02** | -0.039 | -0.019 | -0.015 | *0.006* |

(e) Juvenile cohort PCO axis 9 (from Figure 2.10b)

| PCO contour boundary | > 0.06 | 0.2 to 0.06 | -0.02 to 0.02 | < -0.02 |
|---|---|---|---|---|
| **> 0.06** | *0.123* | 0.064 | 0.069* | 0.047* |
| **0.02 to 0.06** | -0.025 | *0.006* | -0.009 | 0.019* |
| **-0.02 to 0.02** | -0.036 | 0.021 | *0.001* | 0.013* |
| **< -0.02** | -0.006 | -0.026 | -0.021 | *0.000* |

A very similar pattern was displayed by the juvenile cohort (Table 2.4 d,e). While the within group relatedness value was much less in the most southwesterly group ($r$ = 0.046, bound by the spatially interpolated PCO axis 1 contour 0.06), there was still a distinct decay in relatedness between this group and the three other groups, associated with distance along the southwest vector (Table 2.4d). Once again, the pattern displayed significance when the group-wise coancestry coefficient was calculated and tested, with the most southeasterly group expressing significant genetic differentiation from all other groups. Note that the high relatedness seen within the group bound by PCO axis 9 values < 0.06 was only shared between three individuals.

**Evidence of directional gene flow**

Clear evidence for directional gene flow in this population is provided by the comparison of the geographic position of principal coordinate analysis defined contours between both cohorts. A directional spatial shift in genetic composition between the mature and juvenile cohorts is evident when considering PCO axis 1 of the combined mature and juvenile cohort analysis (Figure 2.11). When values of PCO axis 1 of the combined analysis are spatially interpolated for the mature and juvenile cohorts separately, a geographic shift in corresponding PCO axis 1 contours is displayed. Each contour in the juvenile cohort has shifted from its corresponding position in the mature cohort, in a northeasterly direction of around 17 m (Figure 2.11), which is up slope.

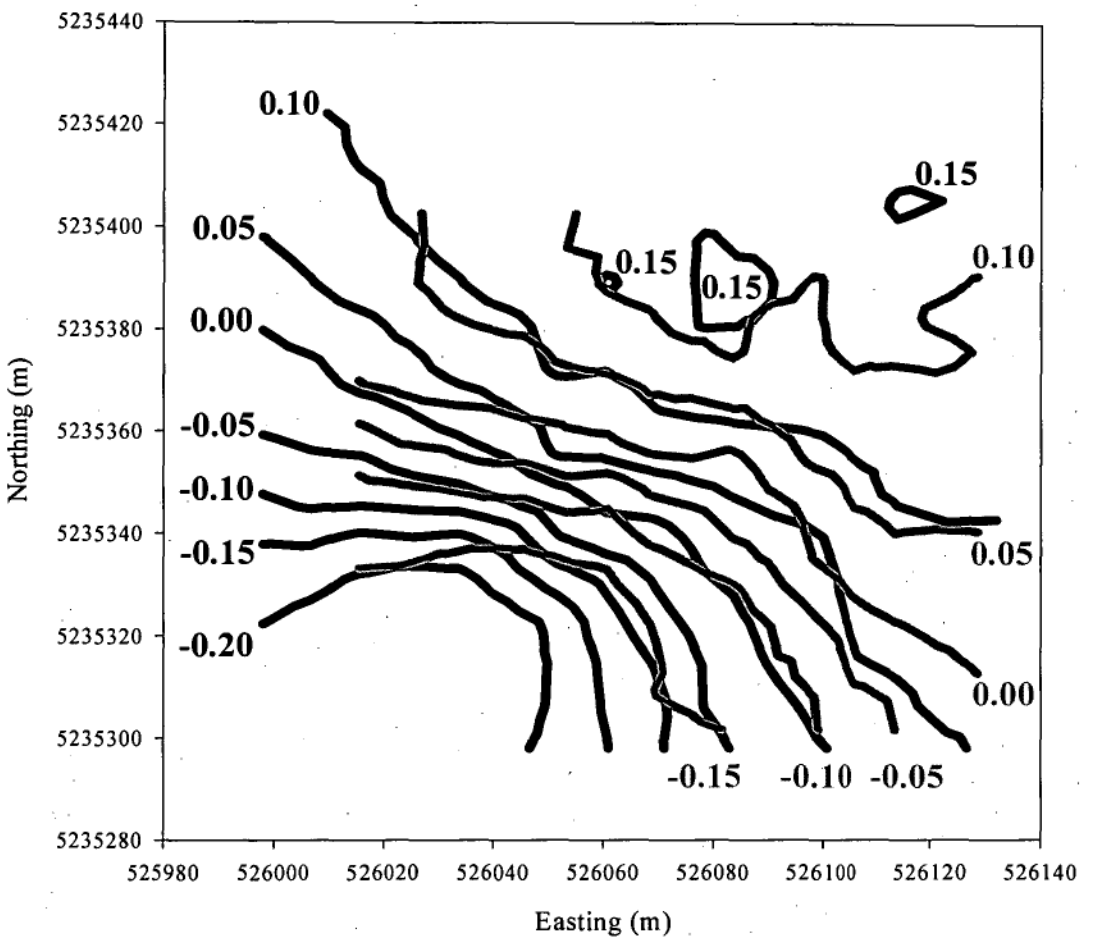



Figure 2.11: A comparison of the spatially interpolated PCO axis 1 contours between the mature and juvenile cohorts. PCO values for each cohort were obtained from the analysis of the combined data from both cohorts with a more extreme smoothing factor than initial interpolations (inverse square; sampling proportion = 0.3; exponent = 1). Hence, contours sharing a similar PCO value are indicative of a similar genetic composition in the two cohorts. By comparing the position of PCO contours between the mature cohort (black) and the juvenile cohort (grey), it is clear that the range of gene frequency combinations have moved around 17 m (average distance shift) towards the northeast in the juvenile cohort.

## Discussion

### Population genetic characteristics

This study included the genotyping of individuals inside a circular sample site of 140 m in diameter (168 mature and 110 juvenile), within a continuous pure stand of native *Eucalyptus globulus*. This is the largest exhaustive sample from a single continuous native population of *Eucalyptus* to be genotyped using microsatellite markers to date and has been highly informative regarding fine-scale within-population patterns of genetic differentiation. The use of recently developed analytical tools has allowed not only the detection of significant spatial genetic structure within the population, but also the identification and characterization of family groups within the sample, their geographic distribution, their level of relatedness and the spatial distribution of genetic change between groups throughout the study site.

This Tinderbox *E. globulus* population displayed marginally higher genetic heterozygosity ($H_e$ = 0.85, $H_o$ = 0.84) than that previously reported for the species in a Tasmania wide study of *E. globulus* ($H_e$ = *0.83* $H_o$ = 0.71) (Jones *et al.* 2002). However, the previous study was based on 37 widely sampled individuals which were genotyped by a slightly less informative array of microsatellite markers (average number of alleles per locus = 12.4, in comparison to 14.6 for the current study). Nevertheless, the current population is highly polymorphic in both the juvenile and mature cohorts, and the microsatellite markers chosen are highly informative.

### Genetic structure and spatial autocorrelation

Well defined spatial structuring of genetic variation in *E. globulus* has been clearly illustrated in the present study, in general concordance with the previous work of Skabo *et al.* (1998), also carried out at a number of sites in the Tinderbox locality. Skabo *et al.* (1998) illustrated significant levels of spatial genetic association in *E. globulus* within distances of 0 to 25 m, and it is within this distance that the most rapid decline in spatial genetic association is displayed in the mature cohort of the present study. However,

while Skabo *et al.* (1998) identified no further local decay in spatial genetic association beyond a distance of 25 m (to at least 500m), the present study suggests that spatial genetic association may extend to at least 55 m between individuals in the mature cohort. In fact, spatial genetic autocorrelation of all individuals, both juvenile and mature, within the current study revealed a basically linear, constant decrease in average genetic association with distance to at least 105 m, which is about the maximum possible distance class comparison within this study (due to the limited number of trees of a greater distance apart within the sample).

While the analysis of the population sample as a whole (including both mature and juvenile cohorts) reveals a continual clinal pattern of decay in spatial genetic association, the potential for the presence of more discrete spatially correlated family groups is suggested by a non-linear decrease in spatial genetic correlation of the mature cohort. As outlined by Vekemans and Hardy (2004) the distance to which spatial genetic structure extends should only be defined when there is a steady decrease in the correlation coefficient until some distance where it stabilizes and shows no further trend. Spatial genetic autocorrelation of the mature cohort (based on the genetic distance measure, Smouse and Peakall 1999) and a similar plot based on the relatedness estimator (Lynch and Ritland 1999) displayed an average distance of between five and 55 m where generally rapid decline in spatial genetic association occurs. This is best displayed when the plot of the relatedness estimator (Lynch and Ritland 1999) is considered, where, after a series of rapid declines and brief plateaux in spatial genetic association, the plot shows no further significant trend beyond 55m. As indicated in Chapter 3 by simulation of various pedigrees based on the mature cohort gene frequencies and the guidelines of Lynch and Ritland (1999), pair-wise relatedness values below zero indicate no significant relatedness. While this is occurring at 47 m and beyond, a mean family group (specifically in terms of recent shared co-ancestry, Bouin 2003), can be realistically estimated as having a radius of around 50 m. In comparison, Skabo *et al* 1998 suggested a smaller family group size of around 25 m in the species. As suggested by Cremer (1966, 1977), seed dispersal in *E. globulus* is likely to be limited to around twice the height of the tree, and it should be noted that with the trees in the mature cohort sharing an average height of 28 m, the estimated family group size in

this case, is very close to that of the distance over which most seed is expected to be dispersed. However, a very gradual residual trend in decay of the spatial association of genetic similarity continuing from 50 m to 105 m (displayed in the spatial autocorrelogram based on genetic distance (Smouse *et al.* 1986)) may be indicative of the more general (and noisy) background of genetic similarity connecting trees over a wider geographic range, suggested by Skabo *et al* (1998).

The plot of mean pair-wise relatedness (Lynch and Ritland 1999) at each five meter distance class of the mature cohort revealed a maximal spatial-genetic association between trees occurring within five to 10 m of each other, expressing a mean relatedness of close to, but less than that of second cousins ($r$ = 0.044). In contrast, Skabo *et al.* (1998) suggest that trees within 25 m of each other may typically exhibit a relatedness of somewhere between that of half-sibling and full-sibling ($r$ = 0.25 to 0.50). When this specific distance class (0 to 25 m) was examined in the present study, using the Lynch and Ritland (1999) relatedness estimator, mean relatedness was even lower ($r$ = 0.02) exhibiting an even greater discrepancy with the study of Skabo *et al.* (1998). It must be noted that the pedigree controls used by Skabo *et al.* (1998) were the result of inter-provenance crosses and the genetic similarity between progeny from these crosses was used as a standard threshold for defining the levels of relatedness within the natural crossing within the Tinderbox locality. The simple matching (SM) coefficient used as a surrogate measure of relatedness in this case does not take into account the inter-provenance variation in population gene frequencies in the control pedigrees. Given that significant genetic variation exists between the various provenances of *E. globulus* (Nesbitt *et al.* 1995; Vaillancourt *et al.* 1995; Dutkowski *et al.* 1997) it is very likely that progeny from the control inter-provenance crosses of Skabo *et al.* exhibited a much lower genetic similarity to progeny of a corresponding pedigree derived from natural crossing within the Tinderbox locality. If this were in fact the case, relationships within the controlled cross pedigree would share a similar simple matching coefficient with relationships within the Tinderbox locality of a significantly more distant pedigree relationship. However, even when this problem is taken into account to some extent, the relatedness between trees within 25 m of each other was still estimated by Skabo *et al.* (1998) at above that of half-siblings.

The trees from the Tinderbox locality used in the *Skabo et al.* (1999) study were sampled from remnant stands within pasture and this may account for this discrepancy, depending on whether or not the trees sampled in that study were founded before or after clearing and hence, progeny derived from natural continuous forest, or from small, genetically isolated patches of parents.

So far, genetic structure in this population has been discussed in terms of simple genetic distance, and more specifically, relatedness. Trees within a patch size of 50 m tend to display a positive level of relatedness and so share recent common ancestry to some degree, in support of the family group cluster pattern (due to limited seed dispersal) suggested to occur generally in native eucalypt forests (Griffin 1990; Eldridge *et al.* 1993; Skabo *et al.* 1998). In addition to limited seed dispersal (Potts and Wiltshire 1997), pollen dispersal in eucalypts is also expected to be restricted to some extent (Hopper and Moran 1981; Savva *et al.* 1988; Griffin 1990) (see Chapter 3) and bi-parental inbreeding depression due to proximity based mating in *E. globulus* has indeed been illustrated by Hardner *et al.* (1998). A significant average spatial association with bi-parental inbreeding and full sib-ship is identified in the present study population, using Lynch and Ritland's (1999) coefficient of fraternity (Trustrum 1961). On average, individuals within zero to five meters of each other displayed mild, but significant bi-parentally inbred relationships and/or full sib-ship, sharing a mean fraternity coefficient value of around 0.028, in comparison to a relatedness coefficient value of around 0.031. The relatedness coefficient is a linear function of 'two gene relatedness' (due to the sharing of a single identical by descent allele at one locus, occurring in simple pedigree relationships) and 'fraternity' ('four gene relatedness', due to the sharing of two alleles identical by descent at one locus, Cotterman 1954, occurring in bi-parentally inbred or full-siblings) (Lynch and Walsh 1998; Lynch and Ritland 1999). Comparison between the spatial plots based on Lynch and Ritland's (1999) mean pair-wise relatedness estimator, and that based on their pair-wise fraternity estimator (Lynch and Ritland 1999) indicates that while, on average, trees within zero to five meters of each other share significant relatedness, this relatedness is likely to be predominantly due to bi-parental inbreeding and/or full-sibship. If these high fraternity values were due solely to full-sibship, the associated relatedness values would be much higher and this was not the

case when each pair-wise relationship was examined (data not shown). Bi-parentally inbred progeny of *E. globulus* display clear inbreeding depression, linearly related to the level of inbreeding (Tilyard *et al.* unpublished). It is therefore possible that inbreeding depression in native populations may not always cause sufficient competitive disadvantage to completely inhibit the survival to maturity of mildly bi-parentally inbred progeny.

**Differences in spatial genetic structure across cohorts**

Variation in the genetic composition of populations between different life-history stages of a species have been highlighted by a number of studies and suggested to be due to a wide range of factors including: selective pressure against inbreeding; method of seed and pollen dispersal; adaptation to different conditions; spatial distribution of individuals; gene flow; founder effects and cohort density (Ritland 1989; Tonsor *et al.* 1993; Epperson and Alvarez Buylla 1997; Aldrich *et al.* 1998; Chung *et al.* 2000; Kalisz *et al.* 2001; Latouche-Halle *et al.* 2003). Most commonly, genetic heterozygosity in forest tree species tends to increase with life history stage (e.g. Latouche-Halle *et al.* 2003, Epperson and Alvarez-Buylla 1997 and Murawski 1995). However, while slight variation in allelic richness (standardized for variation in sample size number, El Mousadik and Petit 1996), between cohorts within the present study was observed (14.5 alleles/locus in the mature cohort versus 13.8 in the juvenile cohort), an $F_{st}$ of 0.000 ± 0.001 (jackknifing over loci) and the absence of significant variation in expected and observed heterozygosity between cohorts indicates no significant variation in genetic diversity between cohorts. The presence of private alleles in the mature cohort (2.9 alleles/locus) would suggest that not all trees are contributing to the juvenile population (see Chapter 3). In addition, private alleles in the juvenile cohort (1.63/locus) suggest a low but apparent level of gene flow from outside the study population. However, given the large number of rare alleles in the population, inheritance of alleles after recombination may result in some alleles not being carried through to the juvenile cohort. This may alone explain the variation between cohorts. In summary these results

suggest that while not entirely panmictic, there is no significant overall reproductive skew between cohorts.

When considering the significant spatial genetic structuring in the mature cohort, the spatially associated bi-parental inbreeding displayed in the mature cohort, and the suggested potential for restricted seed and pollen movement within this population, it would seem reasonable to expect a similar, if not more intensely expressed, spatial genetic autocorrelation in the juvenile cohort of this population (Slatkin 1985; Ennos 1994; Kalisz *et al.* 2001). This has indeed been the case in most life history based studies of spatial genetic structure within forest tree species (e.g. *Pinus pinaster* (Gonzalez-Martinez *et al.* 2002), *Abies balsamea* (Duchesneau and Morin 1999), *Fagus crenata* (Asuka *et al.* 2004), *Cecropia obtusifolia* (Epperson and Alvarez Buylla 1997), *Pinus clausa* (Parker *et al.* 2001), *Platypodium elegans* (Hamrick *et al.* 1993), *Alseis blackiana* (Hamrick *et al.* 1993)). However, as has been observed in a handful of other species, including the forest tree species *Dicorynia guianensis* (Latouche-Halle *et al.* 2003) and the perennial herb *Trillium grandiflorum* (Kalisz *et al.* 2001) and *Plantago lanceolata* (Tonsor *et al.* 1993), spatial genetic structure was less apparent in the juvenile cohort of the present study, in comparison to the mature cohort. Whether genetic distance (Smouse and Peakall 1999) or relatedness (Lynch and Ritland 1999) are considered, the overall gradient in correlation coefficient, degree of mean pair-wise genetic similarity between close individuals and significance of correlation between geographic and genetic distance (Mantels test) was lower in the juvenile cohort than the mature cohort in the present study.

The reduced spatial genetic structure displayed by the juvenile cohort may at first appear somewhat paradoxical, however, there are a number of plausible (albeit speculative) scenarios which take into account the history, ecology and mating system of the study population and when considered, make this situation less surprising:

1. Generation overlap

The higher spatial genetic structure of the mature cohort may be caused by generation overlap within this cohort (Latouche-Halle *et al.* 2003). The real age of trees within the mature cohort is very difficult to determine (Gill 1997) due firstly to indeterminate

growth rates, whereby tree diameter may not correlate well with age. In addition, regeneration of *E. globulus* after fire often occurs via lignotubers originating from the base of burnt stumps (Potts and Wiltshire 1997; Reid *et al.* 1999). Hence, the mature cohort may be composed of a number of generations and comprise potential parent-offspring relationships (Reid *et al.* 1999). In contrast, individuals within the juvenile cohort have not reached reproductive maturity and consequently do not share any parent-offspring relationships with each other (Latouche-Halle *et al.* 2003).

If this were the case however, one would expect an even more clearly defined spatial structure when the mature and juvenile cohorts are analyzed together. The resulting spatial genetic structure from the analysis of the combined juvenile and mature cohort dataset is intermediate to that of the two separate cohort analyses, suggesting that, in this case, generation overlap is probably not a major influence.

2. Limited number of parents for the juvenile cohort

A reduction in the spatial genetic structure of the juvenile cohort may also be due to a very limited number of parents giving rise to that generation. This would result in a juvenile cohort that would lack sufficient genetic diversity to allow maintenance of the spatial genetic structure seen in the mature cohort. However, this is unlikely to be the case as there is no significant difference in genetic diversity between the mature and juvenile cohorts.

3. Change in population density over time

The dilution of spatial structure between the mature and juvenile cohorts may be due to a change in the density of the mature strata of the population over time, from open woodland to closed forest. While an exact history of the vegetation at the Tinderbox Hills does not extend beyond around 100 years, it has been suggested that much of what is now quite closed canopy dry sclerophyll forest in Tasmania has developed from an open woodland habitat (Jackson 1981).

While Hamrick and Nason (1996) point out that when pollen dispersal is random and seed dispersal is localized, significant fine-scale spatial genetic structure may occur, the effect of limited seed dispersal on spatial genetic structure in a population is dependent

on population density (Wright 1943; Hamrick and Nason 1996; Hardner *et al.* 1996; Rousset 1997). In general, open woodland forest tends to display a more pronounced spatial genetic structure (Heywood 1991; Vekemans and Hardy 2004) than closed woodland. While seed dispersal may be limited to around a radius of twice the tree height in *E. globulus* (Cremer 1977), the high density within this population (0.012 trees/$m^2$) means that relatively unrelated progeny in this population have a high likelihood of germinating in close proximity to each other. Indeed, a number of studies indicate that a higher density of adult individuals in forest tree populations may result in a greater overlap of seed shadows and a consequent limitation on the development of genetic structure at the seedling stage (Hamrick *et al.* 1993; Young and Merriam 1994; Hamrick and Nason 1996; Degen *et al.* 2001). The situation in the Tinderbox Hills population may therefore be indicative of a dilution of spatial genetic structure due to an increase in population density over time.

4. Different regeneration regimes

An alternative is that the density of the mature strata has not changed in recent years, but that fewer trees contributed to the recruitment of the mature cohort than to the juvenile cohort. Such an event would occur if a low number of trees had seed crops at the time of a major establishment event such as an intense wild fire, or if the founding population of the mature cohort had been reduced to very few trees after an extremely severe wild fire.

Regeneration of *E. globulus* seedlings may occur after fire, or in response to isolated micro-scale disturbance, such as mature tree death and subsequent light and moisture availability (Duncan 1981). Whilst *E. globulus* is adapted to regenerate rapidly after fire, via lignotubers, epicormic buds and heavy seed dumping (Duncan 1981), survival of individuals is dependent on the intensity of the wild fire (Purdie 1977a; Purdie 1977b; Ashton 1981). However post fire conditions will always result in a fertile ash bed, high light availability and heavy seed dumps from the remaining trees with surviving woody fruit. Whilst some mature trees may regenerate and partially re-close the canopy, there will be patches were progeny from few mothers will be able to survive to reproductive maturity. Parker *et al.* (1996) suggest that in their study of *Pinus clausa* var. *clausa*, the

spatial genetic structure of the juvenile cohort is due to the initial seed dispersal pattern after fire-mediated destruction of the mature cohort and subsequent seedling survival. Indeed, the density of founding trees after fire may have a significant effect on the spatial genetic structure of the post-fire seedling cohort (Knowles *et al.* 1992). However as Boyle *et al* (1990) suggest, the intensity of individual fires and post fire conditions also markedly affect the development of genetic structure. Spatial genetic structure similar to that seen in the mature cohort is more likely to occur under these circumstances than the present mode of regeneration. The fire history of the Tinderbox area supports this hypothesis, with fires generally moving through the area on a twenty year basis. Particularly severe fire has occurred in the area (and over much of southeastern Tasmania) in 1897-98 and 1967 (Wettenhall 2005). Regeneration from the 1967 event particularly, may well have given rise to the individuals that are present in the population today.

Fire has not occurred in the Tinderbox population since 1988 and so regeneration of individuals since then has not been common, with a similar density of juvenile individuals to that of the mature cohort. Seedling recruitment is more prevalent where gaps in the canopy are present. When the annual seed production within the population is considered (Potts and Wiltshire 1997), an extremely low proportion of seeds actually survive to the seedling stage in this situation, let alone to reproductive maturity (only 0.66 juveniles/adult where found in the population). As pointed out by Kalisz *et al.* (2001) even if seeds are not dispersed away from the maternal parent and only one progeny per mother establishes each year (much higher recruitment than what is occurring at the Tinderbox population) the kinship (and relatedness) coefficient of neighboring seedlings (from other maternal parents) would be half that of the parental generation.

5. Micro-site selection

It is possible that an adaptive response to extremely fine scale variation in environmental conditions throughout the site is occurring over time. Under this hypothesis, spatial genetic structure will become more pronounced in particular cohort over time, as the particular individuals that are better adapted to an area within the site

preferentially survive. From an adaptive point of view, despite the initial impression of a relatively homogenous environmental conditions throughout the 140 m wide study site, the effect of micro-scale variation in selective pressure can not be completely discarded (see page 51). The northeasterly side of the study area is near the crest of a hill, and it may be that fine scale adaptation has amplified spatial structure in the mature cohort via selection over time, that has not yet affected the seedling cohort. While a slight change in environmental conditions may to some extent match the clinal genetic variation displayed in both cohorts, the environmental variation is marginal at best, and extremely detailed examination of the micro-site variation on the scale of tens of metres would be required to find any potential correlation.

**Geographic mapping of spatial genetic structure**

While significant trends in spatial genetic structure, quantified in terms of identity by descent and discussed in relation to variation between cohorts, no clear characterization of the actual geographic distribution of particular groups of genetically similar (or family groups), aside from population-wide generalizations, can be drawn from the one dimensional spatial auto-correlation based approach discussed. However, integrated approaches were undertaken in this study to spatially identify actual groups of genetically similar individuals and map their geographic distribution. The first approach taken was the recently developed Bayesian based technique used extensively in various aspects of population genetics, including the resolution of cryptic population structure (Pritchard *et al.* 2000a; Rosenberg *et al.* 2002). This approach indicated that the mature cohort in the present study could not be definitively split into a specific number of distinct groups. Instead, it suggested that the mature cohort could be legitimately split into between three and seven genetically distinct groups. When the individuals were ordinated along two principal coordinate axes, and labeled with the various corresponding structure groups, it became clear that to some extent, the mature cohort consisted of a genetic continuum and that the Bayesian analysis was being forced to split this continuum into specifically delineated groups, hence a similar likelihood being expressed for the three to seven group models. Following the guidelines of Pritchard *et*

*al.* (2000a), the most likely solution of six groups was selected for further analysis, despite the six group model being only marginally greater than the other solutions. The choice of six groups for further analysis was made for illustrative purposes and to provide a reasonable, testable hypothesis, and should not be considered as a definite conclusion as to the number of family groups in the population. This model was also sufficient to clearly illustrate a geographic pattern to the groupings suggested. When the individuals of the mature cohort were mapped and labeled with their corresponding STRUCTURE defined group, the six groups were partitioned (with overlap) in a northeasterly direction. In addition to the autocorrelation analysis carried out in the present study, the geographic clustering of genetically similar STRUCTURE defined groups again supports an isolation by distance model (Wright 1943), however, the groups were partitioned in the same order both genetically and geographically, thus revealing a genetic cline running in an approximately northeasterly direction. Whilst much weaker, a very similar directional pattern was revealed within the juvenile cohort, in agreement with the diluted but still evident spatial structure seen within the juvenile cohort using spatial autocorrelation.

Close examination of the relationship between the six STRUCTURE defined groups and the additional ordination of individuals within the mature cohort suggested that while an underlying clinal pattern in gene frequency change throughout the population exists, another level of complex and even finer scale genetic relationships is present in this population. In other words, while a gradual, directional change in the frequency of a particular group of alleles was initially revealed, the presence of a number of other independent allelic combinations that define additional genetic groupings was indicated. The ordination of the genetic similarity matrix allowed the identification of various principal coordinate axes that illustrate various independent directions in genetic variation throughout the population. As the ordination is based on genetic distance (Smouse and Peakall 1999) between individuals, each principal coordinate axis can be regarded as describing an independent combination of alleles or allele frequencies that can be used to discriminate genetically similar individuals from others within the population. By using an ANOVA, it was possible to partition the relationships between the various STRUCTURE groups in terms of the principal coordinate axes that represent a

particular independent discriminatory group of alleles. The most significant principal coordinate axis partitioned the STRUCTURE groups in the clinal pattern as described, however three STRUCTURE groups were each individually represented by an independent principal coordinate axis. This suggests that there are at least three geographically associated groups of genetically related individuals that differ from the population by sharing an independent and unique combination of alleles. Two of these groups (STRUCTURE groups 1 and 2) are restricted to the southwest corner of the population sample, and overlap significantly. The third group lies between the extreme groups of the cline, geographically situated virtually in the middle of the population sample.

From a methodological point of view, the Bayesian approach presented here has a number of advantages. Most relevant in this case, it allowed the identification of genetically similar groups of individuals, independent of their spatial position and without prior geographically based hypotheses. This means that when individuals were mapped and labeled with their STRUCTURE assigned group, geographic overlap of genetically similar groups of individuals was expressed and did not affect the analysis. This is not the case when using traditional methods that depend on testing spatially delineated subpopulations (e.g. Weir and Cockerham, 1984 and Wright, 1943). However, in terms of illustrating continuous trends in gene frequency change throughout a population, the assignment approach may not be so useful, as this still requires specific predefined genetic boundaries between groups and the subsequent assignment of individuals to one of a limited number of discrete groups. This appears to be the case in the present study.

When principal coordinate axes that displayed a significant spatial autocorrelation were spatially interpolated onto a map of the study population, the above problem was solved and a number of striking patterns of genetic variation were expressed. As already discussed, significant principal coordinate axes are representative of the combined frequency of a particular combination of alleles that best account for variation in genetic distance between individuals (for further detail on this technique, refer to: Cavalli-Sforza 1966, Menozzi *et al.* 1978, Pavesi 2004, Piazza *et al.* 1981, Rendine *et al.* 1999, Romano *et al.* 2003, Shi 1993), and so in essence, allow discrimination between individuals due to various multi-allelic genetic differences. As in a number of studies

(Menozzi *et al.* 1978; Hanotte *et al.* 2002) the spatial interpolations proved a remarkably detailed and effective method for partitioning, summarizing and displaying complex multi-allelic spatial-genetic change. This approach revealed an extremely fine scale spatial-genetic structure in the study population in two geographic dimensions, to a surprisingly high resolution, without assigning specific genetic boundaries. Within the mature cohort, three independent patterns in gene frequency (multi-allelic) change were revealed, that spatially overlapped each other. Firstly, the same southwesterly clinal pattern in genetic change expressed by the STRUCTURE analysis was clearly illustrated. This clinal pattern is characterized by a zone of rapid multi-allelic genetic change (close contours) between two genetic extremes (plateaux) in the southwest and northeast quadrants of the sample distribution. A similar pattern of change along principal coordinate axis 2 is indicative of another independent combination of alleles that illustrate a similar geographic distribution, however in this case the direction of allelic change is more closely partitioned in a purely southerly direction. Comparison of the spatial interpolation of principal coordinate axes 1 and 2 suggests that while the southwest corner of the sample population is inhabited by a group of trees that are clearly genetically different to the rest of the population (in two independent ways, also reflected in the STRUCTURE analysis, groups 1 and 2), the trees in the northwest of the plot also differ genetically to those in the northeast corner of the population. The study site (140 m in diameter) is situated toward the top of an easterly rising hill, with the western side of the plot at the lower side of the rise displaying a reduced slope. The northeastern half of the plot is the steeper part of the study site and does display a slightly drier appearance with less abundant ground cover (approximately 20% less shrub height and abundance). It is possible that the first and second principal coordinate plots are reflecting underlying genetic variation in response to micro-environmental selection, predominantly moisture content and exposure. Indeed, principal coordinate 4 of the mature analysis, which displays genetic similarity between individuals running in a band from the northwestern to southeastern corners of the plot may also indicate a group of individuals adapted to the conditions between the two extremes. However, if the environmental gradient exists, it is subtle and would require detailed assessment to effectively quantify. Suffice to say, the potential spatial ecological pattern is mirrored

by the spatial interpolation of the principal coordinates in both cohorts (with the juvenile cohort expressing similar patterns to those of principal coordinates 1 and 2 of the mature cohort analysis) and the spatial genetic structure does reflect that expected under a model of genetic response to clinal micro-environmental change (Haldane 1948; Fisher 1950; Barber and Jackson 1957; Hanson 1966; Slatkin 1973; May *et al.* 1975).

Whether or not these patterns are due to an extremely fine scale adaptive response to environmental change or purely an outcome of limited gene flow (isolation by distance) is not possible to define without further detailed quantification of environmental variables and testing of the response of genotypes to these factors. However, the patterns expressed by the spatial interpolations are real, tangible patterns in genetic change within a very small area. To quantify the genetic relationships expressed by the spatial interpolation of the principal coordinate axes, the mean pair-wise relatedness (Lynch and Ritland 1999) between and within groups bound by the isopleths (contours) of the interpolation plots was calculated. This revealed a high mean pair-wise relatedness between the 22 individuals (over 10% of the mature cohort) in the most southwesterly corner of the study population. Infact, these individuals share a mean relatedness value (0.106) close to that expected of cousins (0.125, Lynch and Ritland 1999). Keeping in mind that this is the mean relatedness of all the individuals within that area compared to each other, and that a number of individuals share pair-wise relatedness values expected of at least half siblings, it seems appropriate to suggest that these individuals may comprise a tightly geographically clustered family group in the classic sense.

As predicted by the spatial autocorrelation analysis, mean pair-wise relatedness values between groups bound by the isopleths decreased with increasing geographic distance in both cohorts. The additional calculation of the co-ancestry coefficient for each isopleth bound group was undertaken purely to provide an appropriate measure of significance of the genetic difference between groups (Lewis and Zaykin 2001). The matrices of pair-wise comparisons between isopleth bound groups for each spatially interpolated principal coordinate plot in each cohort displayed significant ($P<0.05$) genetic variation between a number of groups, with significance again proportional to the distance between groups. While a significant global association between genetic and geographic

distance was already indicated by autocorrelation, the significance in this case validates the genetic distinctiveness of the groups bound by the isopleths using an all inclusive measure of genetic similarity (co-ancestry coefficient) instead of testing the particular discriminating principal coordinate axis. In summary, from a methodological point of view, this is a way of teasing out trends in genetic variation from a population (ordination), initially independent of spatial information, then introducing a spatial dimension (spatial interpolation), and quantifying and testing the now spatially delineated genetic variation (mean pair-wise relatedness and co-ancestry coefficient). The Bayesian based approach and the spatial interpolation of principal coordinate axes have a significant overlap in outcomes in this study and a very similar, complex pattern of spatial genetic variation in this population is displayed using both approaches.

While this type of spatial genetic structure is most likely due to directional gene flow and other patterns of gene flow reminiscent of isolation by distance models (Wright 1943), it is difficult to make definite inferences as to the causal historical events that gave rise to the spatial genetic architecture evident in this population. However, evidence of directional gene flow is clearly presented when the geographic position of interpolated principal coordinate values are compared between cohorts. The values of the first principal coordinate axis of the ordination of the combined mature and juvenile cohort data were spatially interpolated and separately mapped for each cohort. The mean geographic position of isopleths in the juvenile cohort had shifted around 17 m, up slope, in a northwesterly direction from those of the mature cohort. This may be an adaptive based shift in structure due to a geographic shift in environmental conditions. However, in the absence of a clear environmental gradient, this phenomenon is best described as groups of neutral alleles moving through the population from southwest to northeast, in parallel with the prevailing southwesterly winds. Without moving into direct parentage analysis (Chapter 3), this would appear to be one of the finest levels of directional effective gene-flow ever resolved in a native forest population and provides an insight into the fine-scale processes which underly micro-evolution in these forests.

# Chapter 3: Patterns of local gene flow in *Eucalyptus globulus*

## Introduction

Gene flow via seed and pollen dispersal, in combination with mutation, drift and selection, gives rise to spatially distributed genetic differentiation throughout a species geographic range (Wright 1943; Endler 1973; Levin and Kerster 1974; Hamrick 1982). Limited pollen and seed dispersal alone is sufficient to instigate a patchy distribution of genetic variation within a population (Wright 1931; Adams 1992; Hamrick and Nason 1996), with genetically similar propagules falling in close proximity to each other, and potentially breeding with one another (Wright 1948). On the other hand, long distance pollen dispersal reduces the likelihood of mating between related individuals (Hamrick 1982; Waser and Price 1983). The latter process generally results in less homozygous offspring (Levin 1986) and causes dilution of spatial genetic structure (Ellstrand 1992a) with genetic differentiation among populations inversely related to the number of migrants (or migrant gametes) per generation (Wright 1951; Fenster 1991). In addition, an increase in effective population size (pollen donors) likely to result from wider pollen dispersal will most likely increase genetic variation between siblings within progeny arrays and serve to further dilute spatially associated genetic structure within populations (Levin and Kerster 1974).

The theory of sexual selection in plants (Bateman 1948; Arnold 1994) suggests that while female reproductive success is particularly dependent on the amount of resources that can be allocated to fruit and seed production, male reproductive success is fundamentally dependent on the ability to pollinate as many individuals as possible via widespread pollen dispersal (Smouse *et al.* 1999). Male genetic material is effectively dispersed twice, once via pollen, and again via seed and the pollination system alone is generally accepted as an effective predictor of the distribution of genetic diversity within a population (Hamrick *et al.* 1993), especially where seeds do not display any adaptation to long distance dispersal (Hamrick and Loveless 1986). Consequently, the extent and genetic composition of pollen dispersal in the natural population of *E. globulus,* at the Tinderbox Hills (described in Chapter 2) is of significance when unravelling the

evolutionary processes of the species and may explain the patterns of spatial genetic variation described in Chapter 2.

While gene flow on an evolutionary time scale has been of great interest to evolutionary biologists since the time of Darwin, the quantification of gene flow on an ecological or ongoing time scale has recently grown in importance as a key issue in the assessment and management of threatened populations (Ellstrand 1992b; Caughley 1994; Sork 1998). The result of gene flow on the genetic composition of open pollinated seed lots (in terms of inbreeding and effective population size) has also attracted the attention of tree breeders in response to the requirement of maximising genetic variability and refining predictive breeding models (Hodge *et al.* 1996).

Most theoretical models of gene flow in continuous or fragmented populations have evolved from the original concepts developed by Wright (1943, 1948). Inter-population genetic differentiation was initially estimated using Wright's F-statistics, under various assumptions, allowing indirect estimation of the amount of gene flow from its effect on the distribution of genetic markers (Slatkin 1993; Rousset 1997). A number of the assumptions under this model (e.g. populations characterised by infinite islands with a constant rate of migration between them, Wright 1951) are not biologically reasonable in many cases (Sork *et al.* 1999; Neigel 2002). Development of alternate population models, advances in genotyping ability and an increase in computational power have given rise to a number of potentially more powerful and applicable approaches to studying gene flow in continuous populations (Chakraborty *et al.* 1974; Devlin and Ellstrand 1990; Adams and Birkes 1991; Neigel 1997; Smouse *et al.* 2001; Jones and Ardren 2003; Smouse and Sork 2004).

The most direct method of determining gene flow into a population is via paternity (prior known maternity) or parentage (no known parents) analysis, where as many as possible potential genotyped parents are excluded due to allelic mismatches (Chakraborty *et al.* 1988). If there is more than one potential parent that share the same alleles as the offspring in question, maximum likelihood methods can be used to determine the most likely parent (or parent pair) based on the population frequencies of the shared alleles (Milligan and McMurray 1993; Marshall *et al.* 1998; Gerber *et al.*

2000). Infact, depending on the level of certainty required, the likelihood based approach can be used to assign parents even when allelic mismatches occur (due to genotyping errors, null alleles and mutations between parent and offspring) (Jones and Ardren 2003). The likelihood of mating success of a particular parent can be estimated proportionately over a number of potential parents (Devlin *et al.* 1988) and used to determine a range of reproductive parameters, without necessarily excluding all but one parent (Smouse and Sork 2004). In this case, the possibility of paternity of open pollinated progeny by un-genotyped parents outside the genotyped "neighbourhood" can be included in the analysis (Adams and Birkes 1991; Burczyk *et al.* 2002). The neighbourhood model developed by Adams and Birkes (1991) assumes that the open pollinated progeny may be the result of either self pollination, outcrossing from within a defined neighbourhood of genotyped individuals, or outcrossing from more distant, un-sampled parents (Burczyk *et al.* 2002). Taking into account the mating system (as defined by Neal, 1995) in these relatively simple terms and the likelihood of paternity, reproductive success can be modelled as a function of any potentially predictive variable, using maximum likelihood (Burczyk *et al.* 1996; Smouse *et al.* 1999). This approach provides a framework for appropriate statistical testing of the influence of potential phenotypic characters and pair-wise effects between a male and female on mating success, in combination with likelihood based parentage assignment (Smouse *et al.* 1999).

An alternative approach for estimating pollen movement in particular, is the Twogener model developed by Smouse and Austerlitz (Austerlitz and Smouse 2001; Smouse *et al.* 2001). This method does not rely on the genotyping of potential fathers to infer pollen flow. Instead, it presumes that open pollinated mothers throughout the landscape "sample partially or completely non-overlapping sets of pollen donors" (Smouse and Sork 2004). By relating the degree of overlapping pollination with the distance between mother trees, inference of a number of variables including the pollen dispersal curve and effective population size can be made (Austerlitz and Smouse 2001; Smouse *et al.* 2001; Austerlitz and Smouse 2002; Austerlitz *et al.* 2004).

In a bird and insect pollinated eucalypt species such as *Eucalyptus globulus* (Hingston and Potts 1998a; Hingston and Potts 1998b; Hingston 2002), that has no specialised

adaptation to long distance seed dispersal, seed is expected to fall within two tree heights distance of the mother (Boland *et al.* 1980). The potential for longer distance gene flow over one generation therefore lies in pollen dispersal. As clearly illustrated in Chapter 2, the population of *E. globulus* at the Tinderbox Hills displays remarkable fine scale geographic structuring of genetic variation. Individuals in close proximity to each other (< 55 m) display significant genetic similarity and relatedness, proportional to proximity. The significant geographic partitioning of genetic variation is clearly suggestive of limited gene flow in the population.

The aim of this chapter is to investigate patterns of gene flow within the population discussed in Chapter 2 and define the mating system of *E. globulus* at this site. Of particular significance is the examination of how the spatial structure described in Chapter 2 may influence the amount of bi-parental inbreeding in open pollinated progenies.

## Materials and Methods

### Overview

The identification of pollen donors of open pollinated families allows the subsequent calculation of relatedness between mother and assigned father, and hence provides an estimation of bi-parental inbreeding. Paternity analysis (Gerber *et al.* 2000) was used to assign likely fathers to 555 open pollinated progeny, collected as seed from nine mother trees from near the centre of the study site described in Chapter 2. The 168 mature individuals sampled within an average radius of 63 m from the nine central mother trees were treated as potential fathers. Parentage analysis was also used in an attempt to reconcile the 110 seedlings of the juvenile cohort (of unknown maternity) described in Chapter 2 with their likely parents within the mature cohort. Having assigned likely fathers to the open pollinated progeny where possible, we were then able to manually categorise the mating systems of the nine sampled mother trees. The level of bi-parental inbreeding was estimated by calculating the relatedness (Lynch and Ritland 1999) between mother trees and assigned pollen donors of the open pollinated families. In addition, pair-wise comparisons including the angle and distance between mother and assigned father were made, in addition to estimations of gene flow into the neighbourhood (Devlin and Ellstrand 1990).

The neighbourhood model of Adams and Birkes (1991) allowed the testing of the effects of distance and angle between mother and potential father on mating success. The Twogener (Smouse *et al.* 2001) method was not applicable to this study as the likely pollen flow within the study site was much greater than the distance between the mother trees, a problem that can not be corrected for in this situation (Smouse, P. pers comm.).

**Plant Material and Molecular Methods**

DNA from 168 mature trees and 110 seedlings (within a circular sample site of 140 m in diameter) were extracted and analyzed as described in Chapter 2 (the same genotypes were used). In addition, 555 seeds were collected from nine mother trees, with an average of 62 (59-64) seeds per mother successfully germinated and DNA extracted. These mother trees where situated close to the centre of the study site, with an average radius of 63 m between the mother trees and the edge of the study site (Figure 3.1). One and two year old seed capsules were harvested from four corners of the upper canopy of each mother tree, dried and germinated in vermiculate under laboratory conditions. Families were germinated in separate plastic containers to avoid contamination between trees. Under these conditions, self-pollinated *E. globulus* seed is not expected to have a reduced likelihood of germination in comparison to out-crossed seed (Hardner and Potts 1995; Pound *et al.* 2002; Pound *et al.* 2003).

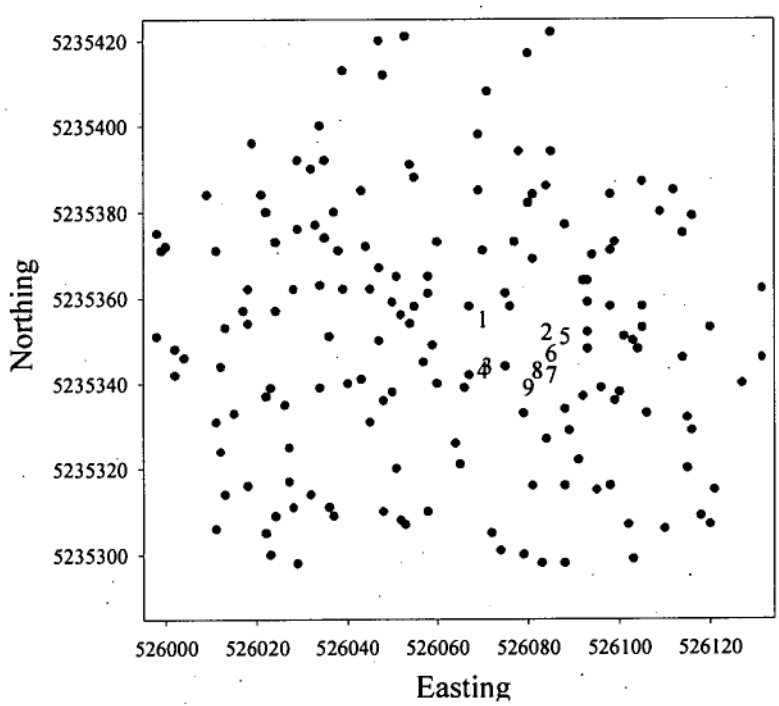



Figure 3.1: Seed from nine mothers within an average or 63 m from the edge of the study were sampled, germinated and fingerprinted with the same microsatellite loci used in the analysis of the mature and juvenile seedlings within the study site (Chapter 2). Numbers 1 to 9 represent tree id numbers 1340, 1347, 1349, 1350, 1357, 1358, 1359, 1360 and 1361 respectively.

The sample site was representative of the *E. globulus* forest dispersed over the Tinderbox Hills. Germinated seedlings were harvested when they were around 20 mm in height and DNA extracted in bulk using the extraction protocol described in Chapter 2, however a bead mill was used for the grinding of tissue. One tungsten carbide bead and an entire seedling (excluding roots) was placed in a single 1.2 ml tube, in racks of 96 and immersed in liquid nitrogen before rapid agitation for two minutes (with a pause at 1 minute to invert racks). This allowed rapid processing of tissue before extraction. Between four and eight microsatellite loci (see Results) from the suite described in Chapter 2 (excluding EMCRC10) were amplified following the protocol provided in Chapter 2.

**Data Analysis**

*Data checking*

In addition to the data checking methods explained in Chapter 2, all seedling genotypes were checked against the maternal genotype at all loci for mistyping with the maternal allele providing an internal standard for every seedling. The loci used in this analysis did not display any evidence of null alleles (Chapter 2).

*Genetic diversity within cohorts*

Genetic diversity between the seed cohort sampled from the nine mother trees, the juvenile cohort and the mature cohort was examined. Observed heterozygosity, expected heterozygosity, private alleles (per individual to standardize between sample sizes), allelic richness (alleles per locus, standardized for sample size, El Mousadik and Petit 1996) and Wright's inbreeding coefficient, $F_{is}$ (Weir and Cockerham 1984)), were calculated for using Fstat V. 2.9.3.2 (Goudet 1995) and GDA V. 1.2 (Lewis and Zaykin 2001). Similarly, these diversity indices were calculated within each of the nine open pollinated families, however allelic richness was not included as the sample sizes of each family were very similar.

*Parentage and paternity analysis.*

The discriminatory power of the loci used in the analysis was quantified by calculating exclusion probabilities for single and the combined loci. Exclusion probabilities provide a measure of the capacity of the loci to validate a particular relationship between individuals by detecting allelic mismatches between offspring and a potential parent or parent pair (Gonzalez-Martinez *et al.* 2003; Jones and Ardren 2003). In this case, three types of exclusion probabilities are calculated. The first is where the capability of the loci to validate paternity, by comparing a mother-offspring pair with a potential father, is tested. The second is where a single potential parent is compared with an offspring that has no initially known parent. The third type of exclusion probability is concerned with comparing a pair of potential parents with an offspring. Each of these three types of exclusion probabilities were calculated following Jamieson and Taylor (1997), using the program FaMoz (Gerber *et al.* 2003). However, the calculation of exclusion probabilities assumes that potential parents or parent pairs make up a random sample from the population, and that the population displays complete panmixis (Double *et al.* 1997). When significant spatial structure is present, as is the case in this population (Chapter 2) and pollen flow is limited, the group of potential pollinators is likely to contain a disproportionate number of relatives (Levin and Kerster 1974). In this situation, the capability of the marker system to exclude all but the real father may be significantly overestimated by exclusion probabilities calculated under the assumption of panmixis (Double *et al.* 1997). In this case a suite of loci that display a very high exclusionary power, may still be unable to exclude all incorrect parents.

Parentage and paternity analysis were carried out with the software FaMoz (Gerber *et al.* 2003), using likelihood ratios (LOD scores), calculated following Meagher and Thompson (1986). In this case, potential parents that share corresponding alleles are initially identified using an exclusion based approach (Marshall *et al.* 1998). The likelihood of paternity/parentage is then calculated based on the population allele frequencies of the alleles that the offspring and assigned parents share (Meagher and Thompson 1986) and used to estimate the most likely parent or parent pair when more than one potential parent or parent pair is initially identified. In addition, simulation of progeny derived firstly, from the real potential genotyped parents (mature cohort

individuals) and secondly, from the mature cohort allele frequencies, was used to predict the probability of an ungenotyped parent from outside the study site providing the same alleles as a genotyped parent from within the mature cohort. This is termed cryptic gene flow (Devlin and Ellstrand 1990; Gerber *et al.* 2003). Comparison of the two simulated arrays allowed the determination of a suitable LOD score threshold that could be used to declare an assigned genotyped parent (or parent pair) as the real parent (or parent pair) as opposed to a possible parent (or parent pair) from outside the stand, with above 95% confidence (Gerber *et al.* 2003). Simulations were repeated 20 times in each case. Despite the spatial structure identified in Chapter 2, the analysis was constrained to using the allele frequencies within the study site to represent those of the trees that surround the site (Gerber *et al.* 2003). Further details regarding the simulation parameters of the particular progeny arrays are discussed in the results. Due to the level of repetition and presence of internal controls during the microsatellite analysis (Chapter 2), a conservative approach was taken throughout the analysis, allowing for zero genotyping error.

An additional constraint can be applied to investigate whether or not the likelihood of the most likely father is sufficiently greater than that of the second most likely father to assign paternity with confidence. In this case, a threshold for the difference between the most likely and second most likely LOD score (delta) can be obtained by again simulating datasets (Marshall *et al.* 1998). For the above simulation of progeny from within the stand, two more datasets were created that provided delta when firstly, the most likely parent was infact the real simulated parent and, secondly, when the most likely parent was the wrong parent. The intersection of the two distributions (data not shown) was used to define a delta threshold to provide confidence that the most likely parent was infact the real parent. However, given that the most likely father in the initial simulations was the real father in above 95% of the cases, the delta threshold was deemed overly conservative and not applied.

*Distance, direction and genetic composition of gene flow.*

The direction of pollen movement from assigned fathers to the mothers of the open pollinated families was firstly investigated by plotting the combined distance of pollinations from a particular direction in a radial plot using Sigmaplot 2000. The position of juvenile seedlings for which parent pairs were successfully assigned (relative to their assigned parents) was also summarized in using a radial plot (Sigmaplot2000). In this case it was not possible to differentiate between male and female members of the assigned parent pair, and so this plot was representative of the directions of the combination of pollen and seed dispersal.

The frequency distribution of distances between assigned fathers and mothers of the open pollinated families were displayed for both the combined nine families and for each family separately. The combined frequency distribution of the distance between all identified crossing pairs (including self pollinations) was compared to the frequency distribution of the distance between the nine open pollinated mothers and all potential fathers (all members of the mature cohort). The distance between assigned mating pairs were manually binned into five and 10 m classes, and various regression models (exponential decay, logarithmic and power) fitted to the frequency distribution using SigmaPlot 2000. The relationship between multiple pollination events between two trees and distance between them was also plotted, with a regression model fitted in the same manner.

The frequency distribution of pair-wise relatedness (Lynch and Ritland 1999) and pair-wise fraternity (Lynch and Ritland 1999) (both discussed in detail in Chapter 2) between the assigned fathers and mothers of the open pollinated families was calculated with the software SPAGEDI (Hardy and Vekemans 2002) and plotted.

Within the open pollinated families, the proportion of self pollinated, half- and full-sibling relationships was estimated using a similar approach to that used by Blouin *et al.* (1996). The percentage frequency distribution of relatedness (Lynch and Ritland 1999) values for three simulated arrays of progeny (unrelated, half- and full-sibs) were plotted together and the intersection of each distribution used to define an upper and/or lower relatedness value threshold for the corresponding relationship. The relatedness

estimator of Lynch and Ritland (1999) will produce a value of one for self pollinated progeny (Lynch and Ritland 1999) and hence, no simulation of self pollinated progeny was necessary. The simulated unrelated, half- and full-sibling relationships were manually created using @RISK 3.5.1 for Excel (1997) based on the real population gene frequencies and generating 61000, 51400 and 50000 individuals respectively, using the eight loci used in the original parentage analysis. Relatedness values were again calculated with SPAGEDI. Within the unrelated progeny array simulation, 2000 unrelated individuals were created, with alleles chosen from the probability distribution created using the mature cohort allele frequencies (assuming independent loci). In this case, the probability of that allele being chosen in the simulated array is equal to its frequency in the real population. Two alleles per locus at eight loci were chosen with replacement, and 61000 pair wise comparisons were made between these unrelated individuals. Pair-wise relatedness was calculated for each comparison using SPAGEDI (Hardy and Vekemans 2002) and plotted as a frequency distribution.

The simulated half-sibling progeny array included 40 families consisting of 50 half-siblings. Each family was based on a parent created from the base allele frequencies as described in the unrelated simulation. 50 half sibs from each parent genotype were then created by selecting one allele from the predetermined parent (mother), and the other allele from the population, determined by mature cohort allelic frequencies. This was repeated 40 times, creating a unique unrelated mother for each family. 51400 pair-wise relatedness comparisons were then made within each family, analyzing one family at a time, using SPAGEDI (Hardy and Vekemans 2002), and including all mature genotypes to reduce bias in allele frequencies.

Similarly, the simulated full-sib progeny array comprised 40 families that consisted of 50 full-sibs each. For each family, two unrelated parents were created from the mature cohort allele frequencies, sampling two alleles per locus. Having created two parents for each family, one allele was chosen from each parent to create a diploid offspring, with all offspring within each family sharing full-sibling relationship. 50000 pair-wise comparisons were made as described above. Mean pair-wise relatedness and standard error for each pedigree was determined from these simulations.

Having defined the range of relatedness values (based on the study population allele frequencies) that corresponded to specific pedigree relationships, it was then possible to categorize the open pollinated progeny into products of self pollination and mating between unrelated, half- and full-sibs, based on the estimated relationship between the mother and assigned father of the progeny. In addition, the number of pollinations from outside the population sample (comprising pollen flow from beyond an average distance of 63 m) were calculated for each open pollinated family.

Variability between families in the frequency of unrelated outcrossing, self pollination and bi-parental inbreeding due to crossing between half- and full-sibs was tested using the $\chi^2$ test for heterogeneity (Rao 1973). The number of progeny from within a family that shared the same father, and are therefore full-sibs, was calculated from the paternity analysis and again variation between families was tested using $\chi^2$ tests for heterogeneity.

Using open pollinated progeny, known maternal and the mature cohort genotypes, multilocus and average single locus outcrossing rates were estimated using MLTR version 2.4 (Ritland 2002). This method does not rely on correct paternity assignment. The analysis was performed using 1000 bootstraps to estimate standard error, using individual progeny as the resampling unit as suggested by Ritland (2002) when family size is less than 12. The proportion of apparent selfing events due to bi-parental inbreeding instead of actual self pollination was calculated by estimating the difference between the multilocus and average single locus out-crossing rate (Ritland 2002). The proportion of full-sibship within the progeny arrays, $r_p$, (or the probability that two individuals drawn at random from an open pollinated family are full-sibs, Ritland 1989) was also calculated using this software. The inverse of this coefficient ($1/ r_p$) can also be used as an estimate of the effective mating population $N_{ep}$, (or the number of males contributing to a family), assuming random mating and independent mating events (Ritland 1989).

*Neighborhood model*

The neighborhood model developed initially by Adams and Birkes (1991) was used to further investigate the mating system of the trees within the study site, with the software NEIGHBOR 2.0 (Burczyk *et al.* 2002). The model defines a neighborhood of a given radius (in this case 63 m) and assumes that progeny arise either from self-fertilization, pollination from a father outside of the neighborhood, or a father from inside the neighborhood. The effect of various parameters on mating success can be tested by the model using maximum likelihood methods (Burczyk and Prat 1997; Burczyk *et al.* 2002). In this case, the distance and angle between the mother and the potential father were tested, assuming identical background allele frequencies to those within the study site (for ease of computation). The interaction between these two variables was also tested. The angle was calculated as degrees from north in a clockwise manner with north being at 0°. The specific model described mating success of a genotyped male within the study population as a function of the level of self pollination in the mother plus the level of pollen flow into the study site from outside plus the distance between mother and father plus angle between mother and father plus the interaction between distance and angle between mother and father. The likelihood functions are defined in Burczyk *et al.* (2002). The mean distance between mothers and all other trees, mean pollen dispersal distance, the mean effective population size $N_{ep}$, self pollination and background pollination (gene flow into the study site) were also investigated and compared to the results of the other methods. However, the requirement of this software for genotypes in the parental population without missing data resulted in the number of mature potential fathers being reduced by 17%. For this reason, the ratio of pollination events due to selfing, crossing within the stand and crossing from outside the stand were manually calculated following paternity assignment as these were more reliable. Likewise, the mean distance between mothers and all other trees and the mean pollen dispersal distance were manually calculated.

## Results

### Information content of loci and heterozygosity within cohorts

The eight loci used in this analysis were extremely informative (Table 3.1a). The combined exclusionary power of the six most informative loci was 1.000 for identifying paternity of the open pollinated progeny that had a previously known mother (Table 3.1b). Additional information is required to identify single parents of the juvenile cohort (with no prior known parent). For this test, the full combination of eight loci provided a potential exclusionary power of 0.998. The combined cumulative exclusionary power of the four most informative loci was 1.000 for identity of parent pairs of seedlings in the juvenile cohort.

Table 3.1: Individual (**a**) and cumulative (**b**) exclusion probabilities of loci are shown for identifying the fathers of the open pollinated progeny (paternity, i.e. one parent known) and the single parents and parent pairs of the individuals within the juvenile cohort (parentage, i.e. no previously known parents). Cumulative exclusionary power is displayed for loci in descending order of information content.

a)

| | Paternity | Parentage | |
|---|---|---|---|
| | | Single | Pair |
| Embra17 | 0.753 | 0.603 | 0.909 |
| Embra10 | 0.749 | 0.597 | 0.905 |
| CRC8 | 0.740 | 0.585 | 0.897 |
| Embra11 | 0.706 | 0.544 | 0.874 |
| Embra12 | 0.692 | 0.527 | 0.863 |
| CRC11 | 0.680 | 0.511 | 0.859 |
| CRC7 | 0.629 | 0.454 | 0.812 |
| CRC2 | 0.580 | 0.399 | 0.770 |

b)

| | Paternity | Parentage | |
|---|---|---|---|
| | | Single | Pair |
| Embra17 | 0.753 | 0.603 | 0.909 |
| Embra10 | 0.938 | 0.840 | 0.991 |
| CRC8 | 0.984 | 0.934 | 0.999 |
| Embra11 | 0.995 | 0.970 | 1.000 |
| Embra12 | 0.999 | 0.986 | 1.000 |
| CRC11 | 1.000 | 0.993 | 1.000 |
| CRC7 | 1.000 | 0.996 | 1.000 |
| CRC2 | 1.000 | 0.998 | 1.000 |

The total number of loci analyzed in each cohort varied depending on the success of PCR amplification and fingerprinting (Table 3.2), with a minimum of four loci still providing sufficient exclusionary power. In this case an exclusion probability for paternity of progeny (harvested seed, with one known parent) using the four least informative loci was still above 0.98 (Table 3.3). Likewise, the exclusion probabilities for single parent and parent pair identification of seedlings in the juvenile cohort (with no previously known parent) were above 0.95 when using the five least informative loci (lowest number of loci successfully amplified in the seedling cohort, Table 3.3). Six individuals in the seedling cohort were excluded from the analysis due to insufficient amplification of loci.

Table 3.2: The full suite of 8 loci was not used to genotype every individual. The number of loci used to genotype individuals within each cohort is summarized.

| | | Number of informative loci | | | | |
|---|---|---|---|---|---|---|
| Cohort | Cohort size | 8 | 7 | 6 | 5 | 4 |
| Mature | 168 | 140 | 19 | 4 | 4 | 1 |
| Juvenile | 110 | 102 | 8 | - | - | - |
| Seedlings | 549 | - | 290 | 40 | 149 | 70 |

Table 3.3: Cumulative exclusion probabilities of the four and five least informative combinations of loci in the analysis were calculated as a baseline to check whether the inclusion of individuals with missing data in the analysis was appropriate. Note that the loci used in this calculation had the least exclusionary power of the 8 loci.

| | Paternity | Single parent | Parent pair |
|---|---|---|---|
| Five least informative loci | 0.995 | 0.965 | 1.000 |
| Four least informative loci | 0.984 | 0.924 | 1.000 |

**Diversity within cohorts**

Surprisingly little variation in expected heterozygosity ($H_e$), observed heterozygosity ($H_o$), and Wright's inbreeding coefficient ($F_{is}$) was observed between the mature, juvenile and seed cohort samples (Table 3.4). While a relatively high number of alleles were detected per locus, the actual allelic richness (average allele number per locus standardized for sample size variation between cohorts) was clearly reduced in the seed cohort in comparison to the mature and juvenile cohort (Table 3.4). There was no major

variation in the diversity indices between the open pollinated families within the seed cohort, except for family1361 which displayed a clearly lower value for the diversity indices and a higher inbreeding coefficient ($F_{is}$) (Table 3.5).

Table 3.4: Genetic diversity indices indicate little variation between the mature, juvenile and seed cohorts. Diversity indices shown are the mean number of alleles of loci ($A$), allelic richness (average number of alleles standardized for variation in sample size) ($R$), expected heterozygosity ($H_e$), observed heterozygosity ($H_o$) and Wrights inbreeding coefficient ($F_{is}$).

| Cohort | $n$ | $A$ | $R$ | $H_e$ | $H_o$ | $F_{is}$ |
|---|---|---|---|---|---|---|
| Mature | 168 | 15.88 | 14.51 | 0.84 | 0.85 | -0.01 |
| Juvenile | 110 | 13.75 | 13.72 | 0.85 | 0.83 | 0.02 |
| Seed | 549 | 14.00 | 11.51 | 0.81 | 0.81 | 0.00 |

Table 3.5: Genetic diversity indices are provided for the nine open pollinated families from which an average of 61.0 seed ($n$) were analyzed. The average number of alleles ($A$), expected heterozygosity ($H_e$), observed heterozygosity ($H_o$), and Wrights inbreeding coefficient ($F_{is}$) are shown for each family. Allelic richness is not included as sample sizes are very similar. Note the reduced level of heterozygosity in seedlings sampled from tree 1361, which displayed a high proportion of inbreeding.

| Sample | $n$ | $A$ | $H_e$ | $H_o$ | $F_{is}$ |
|---|---|---|---|---|---|
| 1340 | 62 | 8.75 | 0.70 | 0.83 | -0.20 |
| 1347 | 58 | 10.14 | 0.76 | 0.87 | -0.14 |
| 1349 | 61 | 8.00 | 0.67 | 0.75 | -0.12 |
| 1350 | 60 | 9.00 | 0.68 | 0.78 | -0.14 |
| 1357 | 64 | 8.71 | 0.69 | 0.83 | -0.20 |
| 1358 | 59 | 9.13 | 0.75 | 0.87 | -0.17 |
| 1359 | 64 | 9.13 | 0.74 | 0.85 | -0.14 |
| 1360 | 59 | 9.25 | 0.74 | 0.81 | -0.11 |
| 1361 | 62 | 7.63 | 0.58 | 0.64 | -0.09 |
| Mean | 61 | 8.86 | 0.70 | 0.80 | -0.15 |

**Parentage analysis of juvenile cohort**

For assignment of single parents to juvenile seedlings with no known parent, simulations suggested that a LOD threshold of 7.5 for single parent identification and 2.0 for parent pairs resulted in an 80.58 ± 0.07% success rate. Twenty simulations were run, based on 10000 simulated progeny and all genotyped individuals in the mature cohort, versus 1000 parents simulated from the mature cohort allele frequency distribution. When these LOD values are applied to the real data, potential single parents for 47 juveniles out of 110 were identified. When the confidence level was further relaxed to 74.96 ±

0.07% (with a LOD threshold of 5.1 for single parents and 0.5 for parent pairs), likely single parents were assigned to 66 individuals in the juvenile cohort. The maximum number of juveniles assigned to one parent was four, with 48 individuals out of 168 from the mature cohort assigned as potential parents. These potential parents showed an even geographic distribution (Figure 3.2).

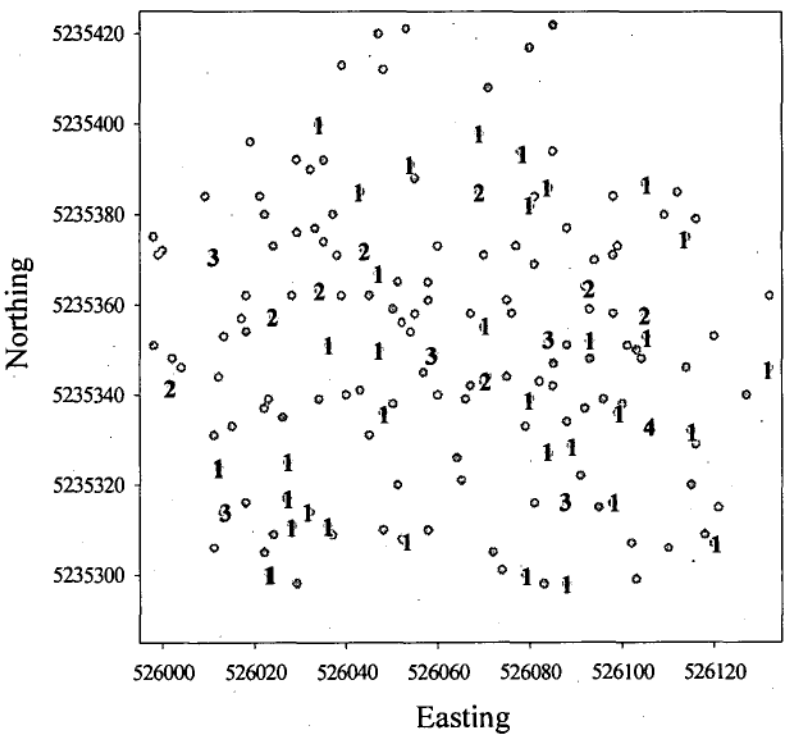



Figure 3.2: Spatial distribution and frequency of parents giving rise to individuals within the juvenile cohort assigned with 75% confidence. Numbers represent position and number of times the individual was assigned parentage to individuals in the juvenile cohort. Grey dots represent the trees in the mature cohort that were not assigned parentage. Juvenile individuals are omitted from the map.

The frequency distribution of the distances between the juvenile seedling and its assigned parent indicates a noisy, but gradual decline in the frequency of parent offspring relationships with distance (Figure 3.3). Comparison to the frequency distribution of the distances between the juveniles for which single parents were assigned, and all possible parents (the mature cohort) indicates a much higher number of parent offspring relationships within 40 m or less of each other, than expected under random seed and pollen dispersal. The gene dispersal curve within the study site was best explained by a negative exponential relationship: number of pollinations $= 14.27^{(-0.027*\text{Distance between trees [m]})}$, with an $R^2$ value = 0.72 ($P < 0.0001$) when the data was binned to 5m distance intervals. However, the regression was only fitted to

individuals more than five meters apart to allow successful fit of the negative exponential relationship.

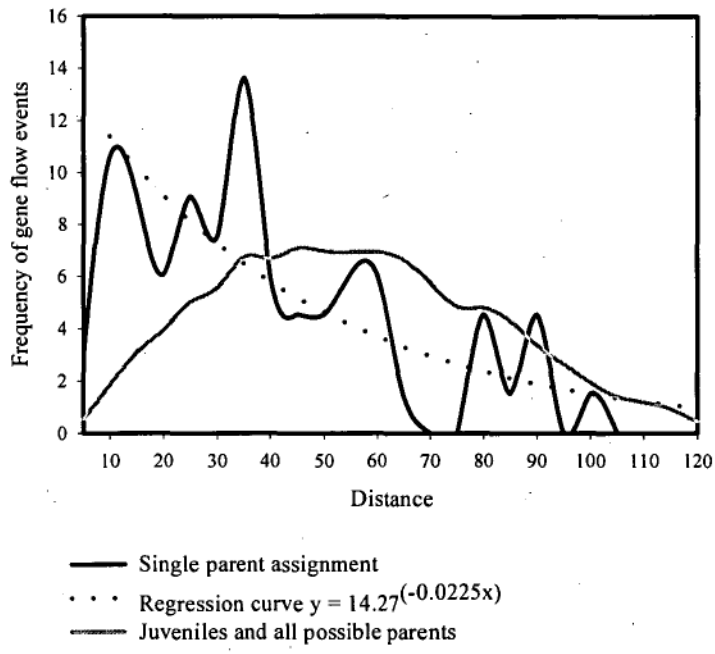



Figure 3.3: The frequency distribution of geographic distance between 66 offspring in the juvenile cohort and their assigned single parent (75% confidence) in the mature cohort. The frequency of parent-offspring relationships between individuals of a particular distance apart are calculated as a proportion of the 66 juvenile individuals that were assigned a single parent. The negative exponential curve $y = 14.27^{(-0.027x)}$ best describes gene flow within this sample site (calculated in SigmaPlot 2000). This curve represents the combined effect of both pollen dispersal to the mother, and then dispersal of the pollinated seed. The grey line represents the frequency distribution of the distances between the 66 juvenile individuals and all the potential parents (168 individuals in the mature cohort).

In the case of parent pair identification of the juvenile cohort, it was necessary to determine a LOD score threshold for both one parent identification and couple identification (Gerber *et al.* 2003). Simulation (10000 simulated progeny, with all individuals in the mature cohort versus 1000 fathers simulated from the mature cohort allele frequency) indicated that a LOD score for single parent identification of 1.0 and for couples of 3.0, resulted in 98.8 ± 0.02% successful parent pair identification over 20 simulation runs. Parent pairs were successfully identified for 18 juveniles under these conditions. The geographic position of each juvenile, relative to its assigned parents, is displayed in Figure 3.4.

The frequency distribution of the distance between the 18 offspring and their assigned parents (data not shown) was almost identical to that derived from the single parent assignment. While the parent pair analysis had a reduced sample size in comparison to the single parent analysis, the confidence of parental pair assignments was significantly higher than single parent assignment (> 95% in comparison to > 75%).

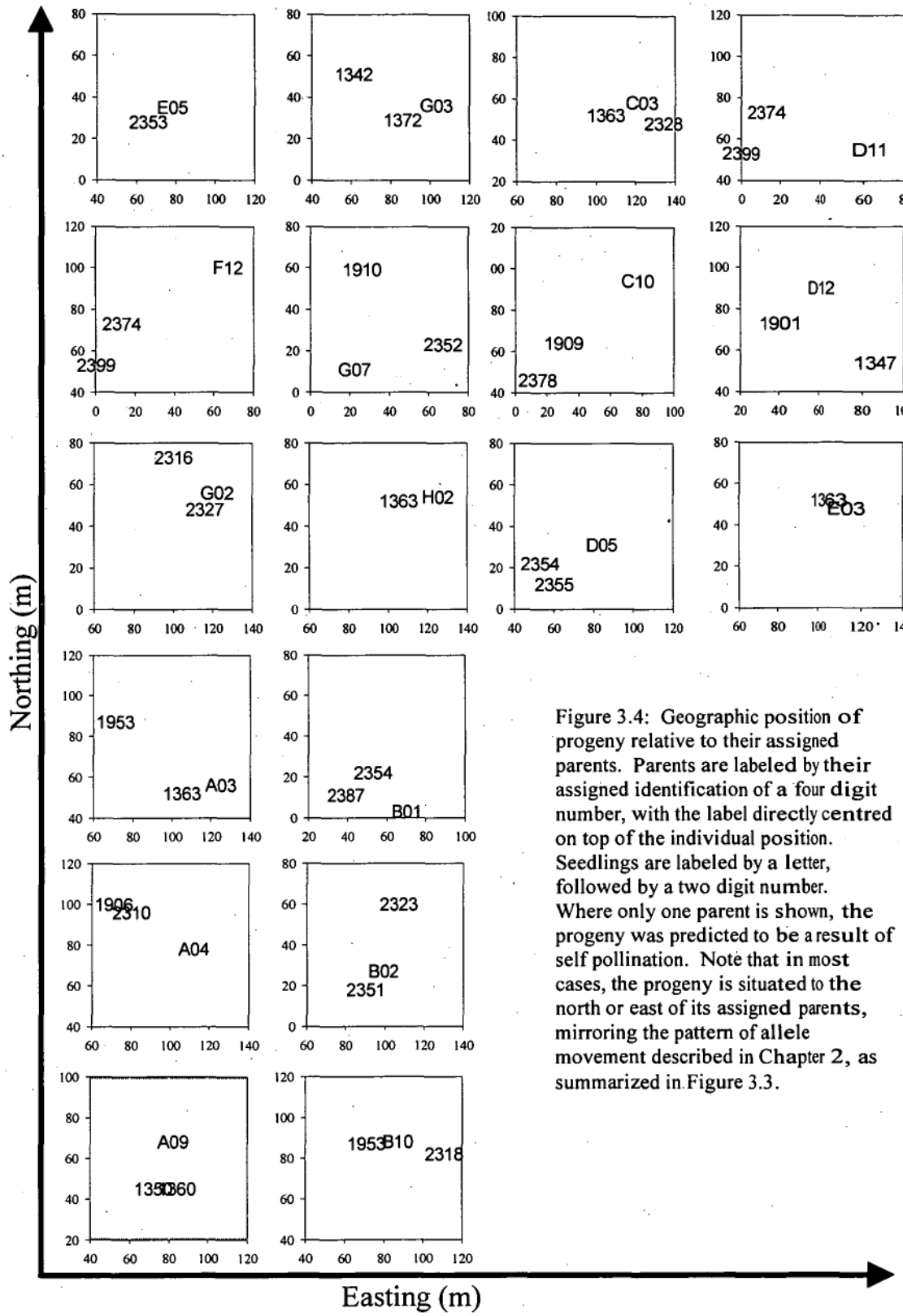



Figure 3.4: Geographic position of progeny relative to their assigned parents. Parents are labeled by their assigned identification of a four digit number, with the label directly centred on top of the individual position. Seedlings are labeled by a letter, followed by a two digit number. Where only one parent is shown, the progeny was predicted to be a result of self pollination. Note that in most cases, the progeny is situated to the north or east of its assigned parents, mirroring the pattern of allele movement described in Chapter 2, as summarized in Figure 3.3.

The direction and distance of seedlings in the juvenile cohort from their assigned parents are summarized in Figure 3.5a. In this case, no distinction can be made between the mother and the father of a particular progeny and the directions of the seedlings from both parents are included. Hence, this distribution displays an overall picture of the direction of movement of genes from the mature cohort to the juvenile cohort (for the 18 seedlings of which parent pair assignment was possible), whether the movement is directly from the mother to the progeny via seed dispersal, or from the father to the mother via pollination and then to the progeny via seed dispersal. A clear easterly trend in gene movement is evident, similar to the northeastern trend in gene movement presented in Chapter 2.

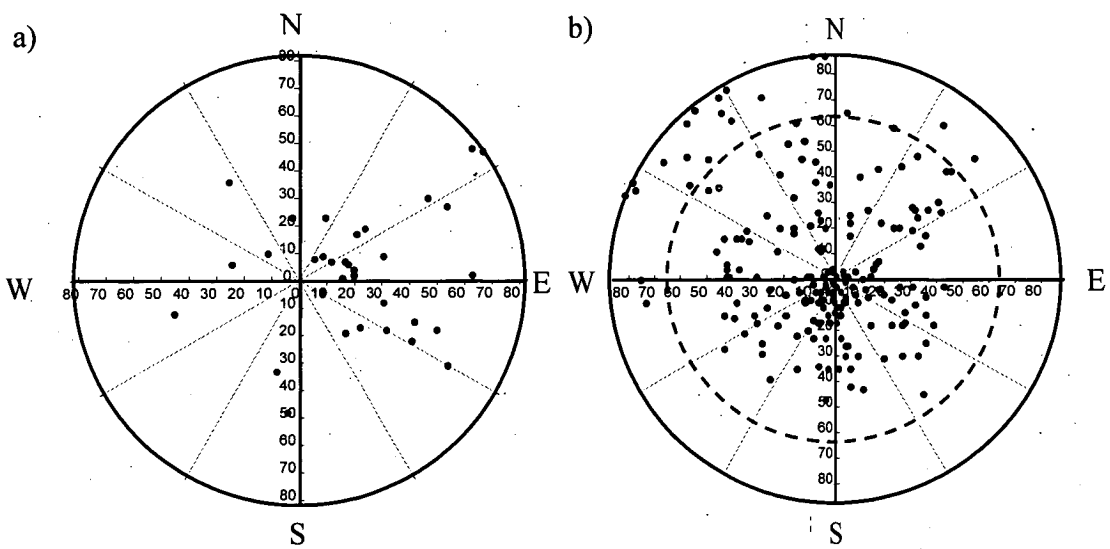



Figure 3.5: The directions and distances between both assigned parents in the mature cohort and offspring in the juvenile cohort are summarized in (**a**), providing a picture of the overall direction and distance of gene flow from mature to juvenile cohorts for the 18 juveniles where parent pair assignment was achieved. All parents are placed at the origin, with the direction and distance of the offspring from the assigned parent represented by a black dot. A clear easterly trend in geneflow is evident. The direction and distance of pollen flow from assigned fathers to the nine females from which open pollinated families were sampled are summarized in (**b**). The 374 fathers are placed at the origin, with the angle and distance between them and the nine mothers of the open-pollinated families indicated by a black dot. The dotted line in (**b**) represents the mean distance between mothers and the edge of the study site. Black dots beyond this line are a result of the offset position (in a southeasterly direction) of the mother trees from the centre of the study site and should be disregarded due to the directional sample bias in this case. When observations within the mean radius of the sample site are considered, no clear directional trend in pollen flow is evident. Note that multiple pollination events between a particular father and mother are not visible in (b), however no directional trend was observed when this was taken into account (data not shown).

**Paternity of open pollinated families**

For paternity analysis of the open pollinated families with a known mother (harvested seed), a LOD threshold of 1.5 resulted in a rate of 95.04 ± 0.06% correct assignments over 20 simulation runs. This was based on the nine real mothers, 10000 simulated progeny and all genotyped individuals in the mature cohort, versus 1000 fathers simulated from the mature cohort allele frequency distribution. With the loci used, cryptic gene flow was estimated at 5.68% (using the same simulations).

Within the real dataset, fathers were assigned to 374 seedlings after application of the LOD threshold of 1.5. In fact potential fathers were actually assigned to 375 progeny based solely on exclusion before application of the LOD threshold of 1.5. Due to the large number of individuals involved, no geographic mapping of individual pollen donors has been presented.

**Composition of open pollinated families**

The average number of assigned fathers per family (minimal effective population size, $N_{ep}$) was 21.2 (Table 3.6) and there was no significant variation between families, tested using the $\chi^2$ test for heterogeneity.

Table 3.6: Number of assigned fathers (and percentage of assigned fathers per progeny). The total number of progeny in each of the nine families are shown in the second column (family size), total = 549. Mean values for number of progeny per family, number of assigned fathers per family and percentage of fathers per number of progeny are indicated in the bottom row.

| Mother id | Family size | Number of fathers | % Fathers |
|---|---|---|---|
| 1340 | 62 | 27 | 43.5 |
| 1347 | 58 | 20 | 34.5 |
| 1349 | 61 | 24 | 39.3 |
| 1350 | 60 | 16 | 26.7 |
| 1357 | 64 | 18 | 28.1 |
| 1358 | 59 | 20 | 33.9 |
| 1359 | 64 | 28 | 43.8 |
| 1360 | 59 | 18 | 30.5 |
| 1361 | 62 | 20 | 32.3 |
| mean | 61.0 | 21.2 | 34.4 |

The analysis using the correlated paternity approach of Ritland (2002) estimated a mean effective population of 19.23 pollen donors per family, made without prior knowledge of potential paternal genotypes. The Neighborhood model approach of Adams and Birkes (1991), using the software NEIGHBOR 2.0 (Burczyk *et al.* 2002), suggested a slightly larger effective population size of 23.39 (19.53% of the genotyped potential parents) within an average radius of 63 m of the mother, not taking into account pollinations from further away.

The paternity analysis suggested that 10% of the 549 progeny were the result of self pollination (Table 3.7), matching the estimates of the Neighborhood model analysis (selfing = 10.0%, Std. dev. = 1.2%) and the correlated paternity approach (selfing = 10.0 ± 4.5%). Comparison between multilocus and single locus outcrossing rates (Ritland 2002) allowed the proportion of assigned self pollinations that are actually due to bi-parental inbreeding to be estimated at 1.6 ± 1.4%, indicating a very low likelihood that this is actually occurring in this study and that selfing rates do represent true self pollination events. The average proportion of individuals within a family that shared a common father (full-siblings) with at least one other seedling was at least 32% (as calculated manually by comparing paternity assignments) and this varied significantly ($\chi^2_{df} = 8$; $P < 0.001$) between families (Table 3.7). However, the probability that two individuals randomly chosen from one family are full-siblings was estimated as 0.053 ± 0.008, using Ritland's (2002) correlated paternity approach with the software MLTR (Ritland 2002).

Table 3.7: The average number of the three relationships within the open- pollinated families are presented. Significant variation between families was tested using the $\chi^2$ test for heterogeneity (Rao 1973).

| Relationship | Percentage | Inter-family variation |
|---|---|---|
| Self pollinated | 10% | *** |
| Full-sibling | 32% | *** |
| Half-sibling | 58% | ** |

**Pollen dispersal**

The proportion of background pollination (pollen flow from outside the study population) was manually estimated as 33% from the paternity assignment approach. The combined frequency distribution of the distances between the nine mothers and the assigned fathers of their open pollinated progeny (including self pollination) showed a rapid decline in the frequency of pollinations with an increasing distance between members of the mating pair (Figure 3.6). This decline was more rapid than that expressed by the overall gene dispersal frequency distribution which is a combination of the effect of both seed and pollen movement (Figure 3.3). The overall gene dispersal curve shows a relatively higher frequency of dispersal events between 20 and 40 m than the pollen dispersal curve alone, suggesting that seed dispersal between 20 and 40 m is very common in the site. The pollen dispersal curve for each individual family is shown in Figure 3.6, indicating clear variation between families in the spatial source of pollen.

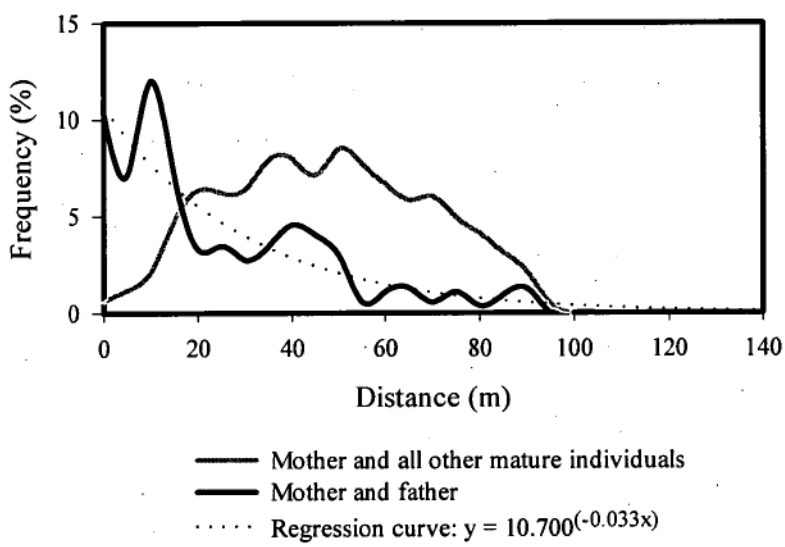



Figure 3.6: The frequency distribution of geographic distance between mother and assigned father for 374 open pollinated progeny sampled from nine mother trees. The frequency of pollinations between parents that are a particular distance apart are calculated as a proportion of all 549 progeny, including the 33% of pollinations that came from outside the sample population. The negative exponential curve $y = 10.700^{(-0.033x)}$ that best describes the local pollen dispersal curve is displayed.

The local pollen dispersal curve when the data was binned to 5m distance intervals was best explained by a negative exponential relationship, with the estimated number of pollinations = $10.700^{(-0.033*\text{Distance between trees [m]})}$, with an $R^2$ value = 0.86 ($P < 0.0001$) (calculated in SigmaPlot 2000) (Figure 3.6). Binning to 10 m distance intervals allowed a slightly better fit of the regression curve ($R^2 = 0.90$) for the negative exponential relationship ($y=43.638^{(-0.036x)}$), however the power function $y = 281.026x^{(-0.923)}$ fitted even more closely with an $R^2$ value = 0.92. Sigmaplot was not able to fit a power

function regression curve to the data when binned to 5m intervals. This comparison indicates that care should be taken when interpreting a pollen dispersal curve calculated with this approach, as the slope and intercept of the regression is evidently biased by the method of binning distance classes prior to the regression analysis.

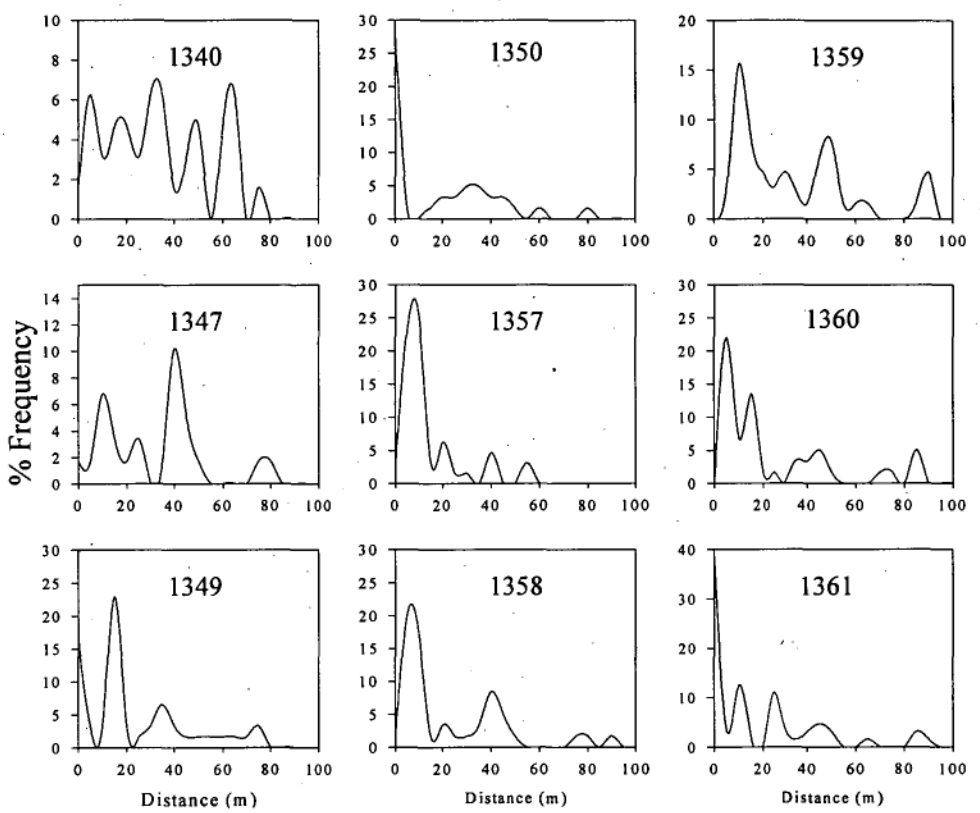



Figure 3.7: The frequency distribution of distances between each sampled mother (tree id number shown in each graph) and assigned fathers of the corresponding open pollinated family. Each graph represents one of the nine open pollinated families examined, the combined data for which is shown in Figure 3.5.

It is clear that pollination between trees separated by 15 m or less is much more likely to occur than between trees separated by a greater distance. 37% and 61% of all progeny were assigned to parents separated by 15 m and 50 m or less respectively (including self pollination events). 33% of pollinations occurred due to pollen traveling from outside the sampled population (excluding the three ungenotyped trees within the site) and so, the complete frequency distribution of all pollinations no doubt displays a much longer tail. A mean pollen dispersal distance within the study site (including self pollination

with distance defined as zero and excluding pollinations from outside of the stand) of 22.14 m (range 1.4 m to 84 m) was estimated manually by calculating the mean distance between mothers of the open pollinated progeny and the assigned fathers. In comparison, mean pollen dispersal was also calculated as 13.73 m using the NEIGHBOR 2.0 software (Burczyk *et al.* 2002), possibly affected by the exclusion of individual genotypes that contained missing data. The Neighborhood model allowed the testing of the effect of distance and direction between mother and potential father on the likelihood of pollination success. The effect of distance between trees on the mating success of the father was estimated as -0.103 (std dev = 0.008) indicating a significant exponential decrease in mating success with increasing distance from the mother (Burczyk *et al.* 2002), reflecting the pollen dispersal curve. The angle between mother and potential father displayed no significant effect on mating success. This is reflected in the plot of the angle and distance between the assigned fathers of the open pollinated families and the corresponding nine mothers (Figure 3.5b), where no asymmetry in pollen dispersal is evident. There was no significant influence on mating success of the interaction between angle and distance between parents.

A striking negative exponential relationship between the number of progeny within a father sires per father and mean distance between the parents was revealed when the two variables were plotted together (Figure 3.8). The distribution best fitted a power function with an $R^2 = 0.83$ ($P < 0.001$). On average, full-siblings are much more likely to be the product of mating between trees within close proximity to each other, and the number of successful pollinations between a mating pair appears to be proportional to the proximity of the pair.

**Simulated progeny arrays**

To quantify bi-parental inbreeding in the nine open pollinated families, the relationship between each mother and assigned father of the progeny was determined. This was done by comparing the relatedness values between the assigned mating pairs with the defined relatedness thresholds determined from the simulated full-sibling, half-sibling and

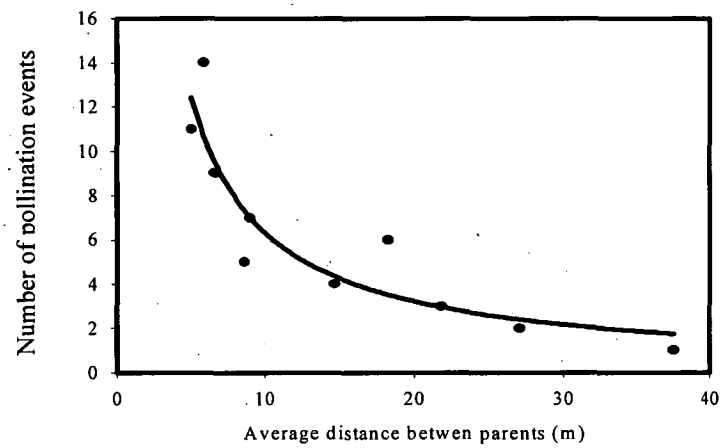



Figure 3.8: A power function with negative exponential ($y=59.45x^{(-0.97)}$) best describes the relationship between the number of pollinations within a family due to a particular father and the average distance between the mating pair.

unrelated individual datasets. When no father was assigned, it was assumed that the progeny was a product of pollination from a father outside the study population. Due to the spatial structure displayed in Chapter 2, these pollinations were assumed to be between unrelated individuals. For each simulated dataset, the mean pair-wise relatedness values for the simulated groups were mildly lower than that expected from Lynch and Ritland (1999) (Table 3.8). The distribution of the simulated full-sibling group appeared relatively normal, with the half-sibling and unrelated datasets displaying mild skew with low frequency tails extending towards the higher values (Figure 3.9).

Table 3.8: Relationship, number of individuals within the array, observed and expected mean pairwise relatedness values (Lynch and Ritland 1999) for simulated progeny arrays, based on mature cohort gene frequencies.

| Simulated relationship | Number of simulated individuals | Mean pair-wise relatedness ± s.e. | Expected mean pair wise relatedness |
|---|---|---|---|
| Unrelated | 61000 | -0.003 ± 0.000 | 0.000 |
| Half-sibling | 51400 | 0.219 ± 0.001 | 0.250 |
| Full-sibling | 50000 | 0.481 ± 0.001 | 0.500 |

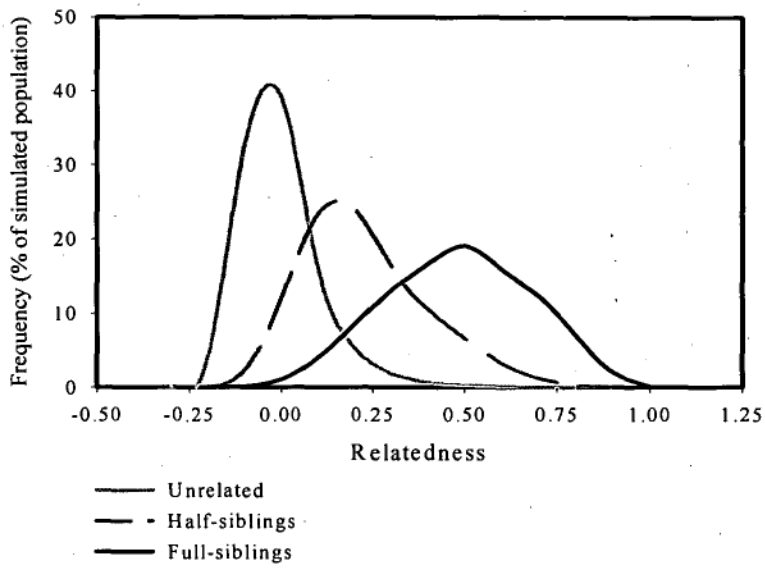



Figure 3.9: Percentage frequency distribution of the pair-wise relatedness value between comparisons made with the simulated unrelated, half-sibling and full-sibling array. The point at which each distribution intersects the other is used as a threshold for relating relatedness values to a particular pedigree relationship.

The range of relatedness values corresponding to self pollination and mating between full-siblings half-sibling and un-related individuals were defined as 1.00, 0.34 – 0.99, 0.08 – 0.34, and < 0.08 respectively (Figure 3.9). The expected error when classifying the relationship between a mating pair based on these thresholds was estimated by calculating the proportion of each simulated frequency distribution that lay beyond the defined threshold. 24.5% of the simulated full siblings displayed a relatedness value below the threshold of 0.34 for full siblings. However 23.2% of the simulated half siblings displayed a relatedness value above the upper threshold of 0.34 for half siblings. This indicates that the likelihood of misclassifying a full sibling relationship as a half sibling relationship and conversely, a half sibling relationship as a full sibling relationship is almost equal, based on this defined threshold. While this reduces the confidence in the classification of a particular individual relationship, a similar chance of misclassification in both directions suggests that over a large population sample, the mean classifications in this case are likely to be reasonably unbiased. The proportion of simulated half siblings that displayed a relatedness value below the lower threshold of 0.08 was 23.18%, in comparison to the 17.6% of the simulated unrelated individuals that displayed a relatedness value above the same threshold. This indicates that the defined threshold for differentiation between half sibling and unrelated relationships may be slightly biased towards under estimation of the number of matings between half siblings in the population.

**Bi-parental inbreeding**

When the frequency of pair-wise relatedness values (Lynch and Ritland 1999) between all assigned mating pairs were displayed (Figure 3.10), a disproportionate frequency of matings between individuals sharing a relatedness value of around 0.21 equivalent to that displayed by the simulated half siblings was evident. This peak is not expected under random mating when the frequency distribution of pair-wise relatedness values between the nine mothers and all trees within the mature cohort is considered (Figure 3.10). 43% of the progeny were assigned to fathers within the study site that did not share any close relationship with the mother (Table 3.9). 13% of the progeny had parents that shared relatedness values within that expected of half-sibling relationships and one percent of the progeny were the likely product of crossing between full-siblings (Table 3.9).

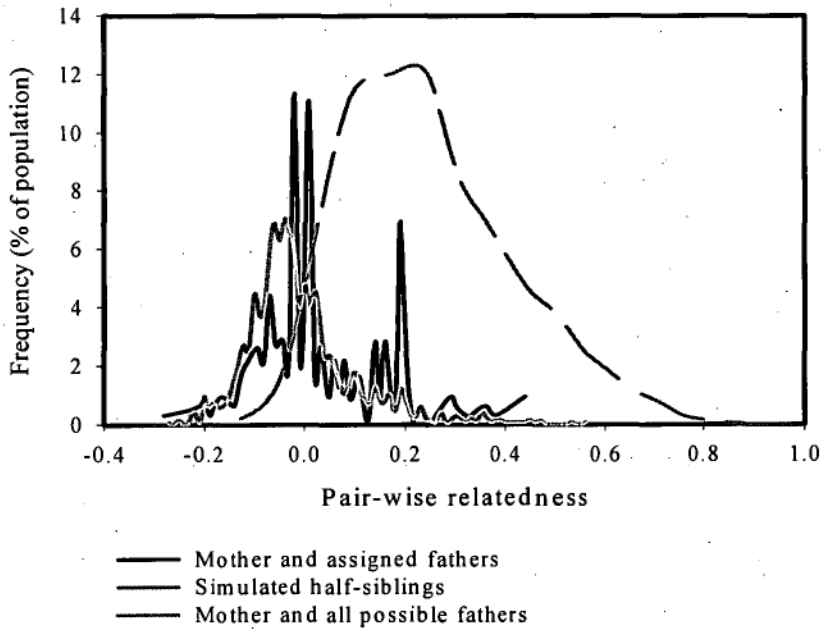



Figure 3.10: The frequency distribution of pair-wise relatedness values between individuals within the simulated half sibling array, in comparison to the frequency distribution of the pair-wise relatedness values of all identified mating parental pairs (excluding self pollination events) of the nine open pollinated families. Note the increased frequency of matings between individuals sharing a relatedness value of around 0.20, corresponding to a half sibling relationship.

Significant heterogeneity was observed between females in the level of self pollination, crossing between half-siblings, crossing with un-related trees within the study site and long distance pollinations (> 62.5 m). This is clearly evident when comparing the nine individual frequency distributions of pair-wise relatedness between assigned parents of the nine open-pollinated families (Figure 3.11).

Table 3.9: The frequency (combined across the nine open pollinated families) of self pollination events and outcrosses with fathers that were likely full-siblings, half-siblings, unrelated (within the study population), and unrelated (outside the study population). This was estimated from the number of mating events between individuals that shared a relatedness value corresponding to a particular relationship determined by simulation. Nine open pollinated families were examined with a total of 555 progeny, with 374 progeny assigned to fathers within the study site. Significant variation in the number of pollinations due to crossing between individuals sharing the particular relationship between families ($\chi^2$ test) are shown (df = 8, *** = p < 0.001). ¥No assigned fathers, indicating pollen flow from outside the study population, and presumed un-related matings (based on spatial distribution of relatedness, Chapter 2).

| Relationship | Relatedness value range | Frequency | Inter-family range (%) |
|---|---|---|---|
| Self | 1.00 | 10% | (0-40)*** |
| Full-sibling | 0.34 – 0.99 | 1% | (0-3)n.s. |
| Half-sibling | 0.08 – 0.34 | 13% | (2-24)*** |
| Un-related | < 0.08 | 43% | (16-62)*** |
| Outside sample¥ | - | 33% | (10-56)*** |

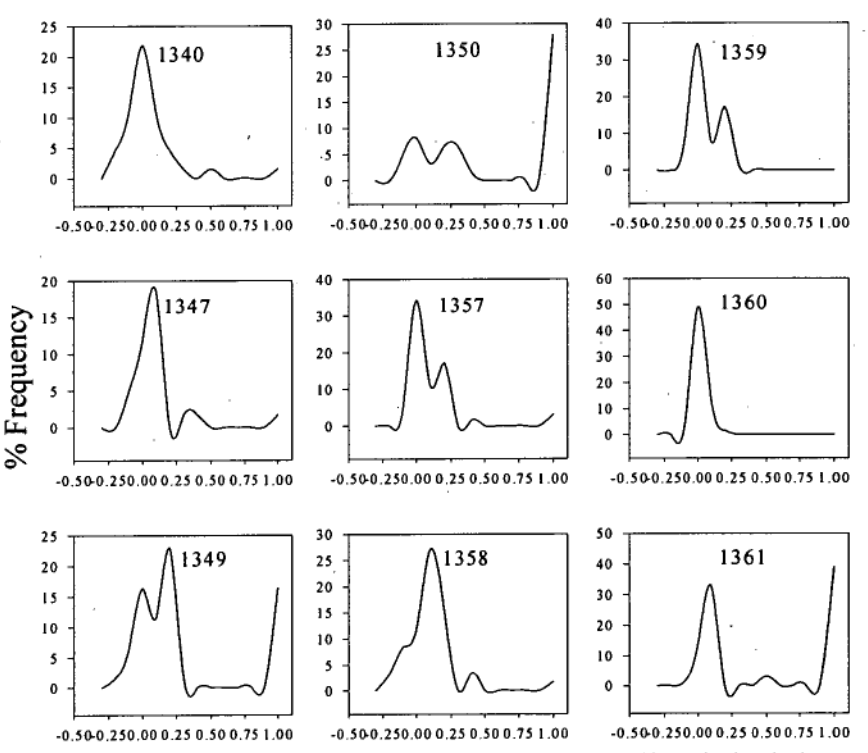



Figure 3.11: The relative frequency distribution of pairwise relatedness for each sampled mother (tree id number shown in each graph) and assigned fathers of the corresponding open pollinated family. Each graphs represents one of the nine open pollinated families examined, the combined data for which is shown in Figure 3.10. A relatedness value of 1.00 is due to self pollination events in this case.

The frequency distribution of pair-wise fraternity values (Lynch and Ritland 1999) between all assigned mating pairs displayed a slightly skewed tail (Figure 3.12) above that expected under random mating (as suggested by the plot of the pair-wise fraternity values between the nine mothers and all other mature individuals). This slight skew towards high fraternity values may suggest a low level of bi-parental inbreeding between the grand parents of the progeny. Variation between families was evident in the frequency distribution of pair wise fraternity values (Figure 3.13, f<0.5).

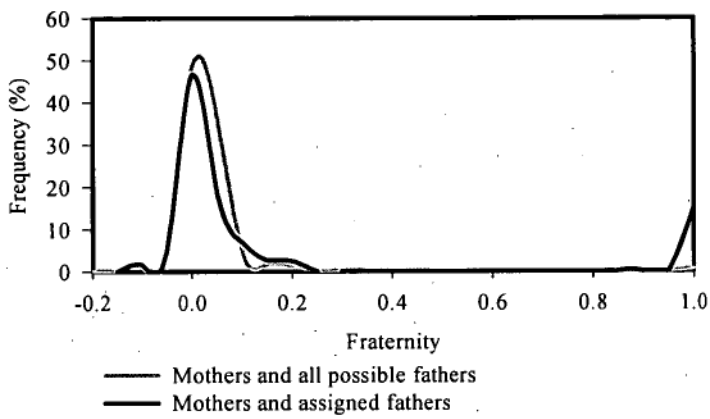



Figure 3.12: The frequency distribution of pair-wise fraternity between the assigned parents of all the open pollinated progeny is compared to that between the nine mothers and all possible parents. Comparison of the two curves indicate that fraternity values above 0.1 are slightly more frequent between the mating pairs, than expected under random mating. Pair-wise fraternity values of 1.0 represent self pollination events.

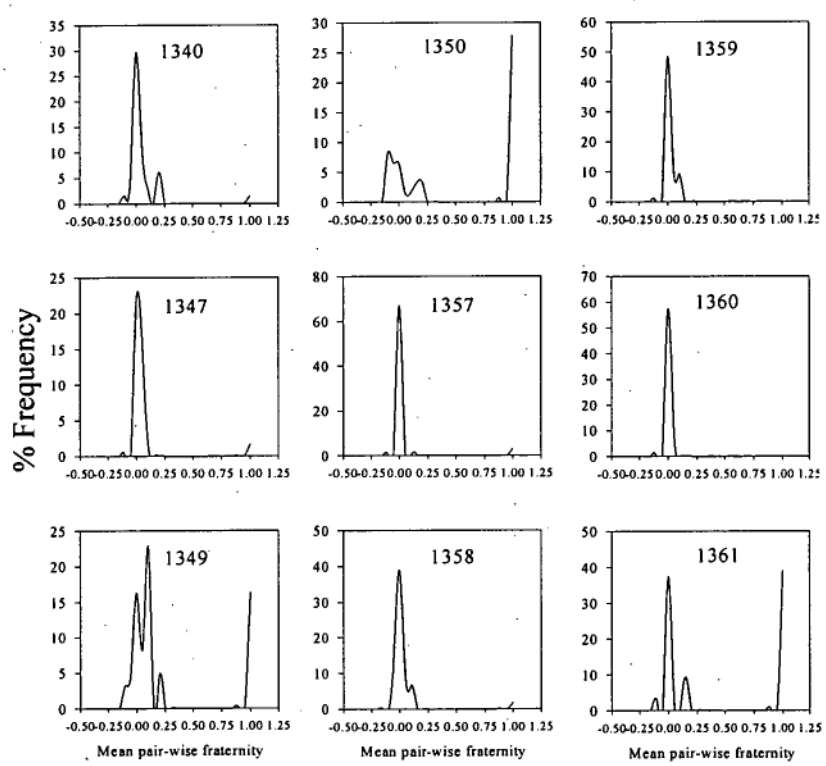



Figure 3.13: The frequency distribution of pairwise fraternity values between each sampled mother (tree id number shown in each graph) and assigned fathers of the corresponding open pollinated family. Each graph represents one of the nine open pollinated families examined, the combined data for which is shown in Figure 3.12. A pairwise fraternity value of 1.00 is due to self pollination events.

Data from Tilyard *et al.* (un published) indicates that a linear reduction in average growth of families occurs with the severity of inbreeding (Figure 3.14). The effect on growth of a particular level of inbreeding can be combined with the amount of the corresponding level of inbreeding that is occurring in the current study population to provide an estimate of the average reduction in growth of open pollinated families due to both self pollination and bi-parental inbreeding in native populations of *E. globulus* (Table 3.10).

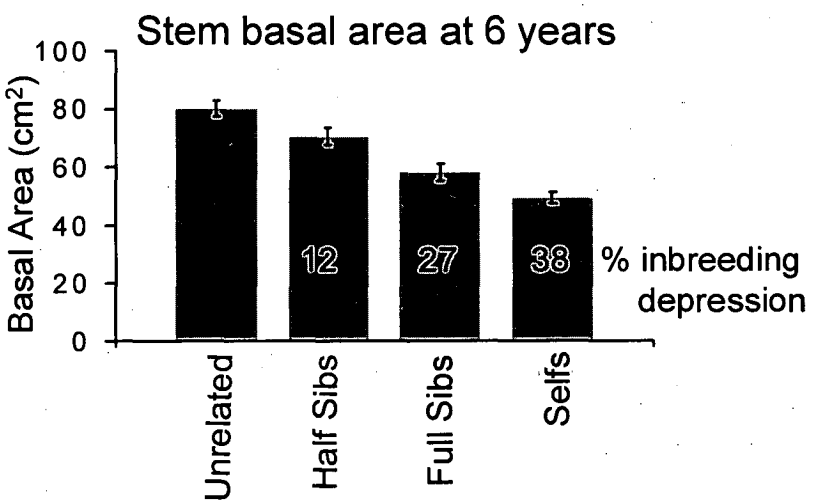



Figure 3.14: A linear reduction in average growth of controlled cross *E. globulus* families is seen, in proportion to the level of relatedness between parents. The data is based on stem basal area after 6 years and clearly shows the effect of bi-parental inbreeding on growth (P. Tilyard, C. Hardner and B. Potts unpublished data).

While self pollination was estimated to reduce the average growth (basal area at six years) of the progeny by an average of 3.8%, the additional effect of mating between full siblings (0.27%) and mating between half siblings (1.56%) is expected to result in a total effect of inbreeding depression reducing average stem basal area growth of an open pollinated family by 5.63% after six years.

| Level of inbreeding | Average reduction in expected growth (%) |
|---|---|
| Self pollination | 3.80 |
| Full sibling | 0.27 |
| Half sibling | 1.56 |
| Total | 5.63 |

Table 3.10: The estimated average reduction in stem basal area of an open pollinated family of *E. globulus* in the natural environment due to self pollination and bi-parental inbreeding between half and full siblings.

## Discussion

As the significant fine-scale spatial genetic variation described in Chapter 2 would suggest, the Tinderbox Hills population of *Eucalyptus globulus* displayed spatially limited gene flow. In concordance with the accepted leptokurtic pollen dispersal curve expected in most seed plants (Levin and Kerster 1974), including eucalypts (Barber 1965), the likelihood of pollinations within the nine open pollinated families decreased exponentially with distance between the potential mating pair (within the area of the study site). This pattern has been displayed in other eucalypt species including *E. nitens* (Barbour *et al.* 2005) and *E. regnans* (Burczyk *et al.* 2002). The effective pollen dispersal (frequency distribution of pollinations versus distance between mating pair) was strikingly different to that expected under random mating, and the application of the Neighborhood model (Burczyk *et al.* 2002) confirmed a significant exponential decrease in the likelihood of successful pollination with distance between individuals. The relationship between likelihood of pollination and distance in the present study was best summarized by a negative exponential function, with the estimated number of pollinations = $10.700^{(-0.033*\text{Distance between trees [m]})}$ (binning to five meter intervals, $R^2$ value = 0.86, P <0.0001). It is, however, interesting to note that the pollen dispersal curve of the present study fitted more closely to a power function with a negative exponential when the data was binned to 10 m intervals, which was very similar to pattern described for *E. nitens* (Barbour *et al.* 2005). The binning approach used in the present study was different to that of Barbour *et al.* (2005) who were studying pollen dispersal on a wider scale, at a different site and in a different species.

The average pollen dispersal distance within the current site as estimated from paternity assignment was 22 m (ranging from 1.7 m to 84 m). Ten percent of the open pollinated progeny appeared to be due to self pollination. 29% of all the unselfed progeny examined were pollinated by trees that were 15 m or less from the mother tree. 36% of all unselfed progeny (33% of all progeny) were pollinated by fathers further away than around 63 m from the mother trees. Cryptic gene flow occurs when gene flow from outside the genotyped neighborhood is mistakenly assigned to pollen donors within the neighborhood that share the same alleles as the real un-genotyped father (Devlin and

Ellstrand 1990). The risk of this occurring is directly proportional to the information content of the markers used, and if unaccounted for, may result in an underestimation of gene flow into the study site (Sork *et al.* 1999). In this study, cryptic gene flow was estimated at 5.68% using a simulation approach (Gerber *et al.* 2003), and at most would increase the proportion of pollinations from outside the population sample from the estimated 33% to 39%. However, this is likely to be countered by possible contributions from now dead trees and the three ungenotyped individuals within the study site (as explained in Chapter 2).

Insects including the honey bee *Apis mellifera,* the bumble bee *Bombus terristrus,* the native bees such as *Hylaeus honestus* and the beetles *Phyllotocus rufipennis* and *Mordellistena* spp. play a role in pollination of *E. globulus* (Hingston and Potts 1998a; Hingston *et al.* 2004a; Hingston *et al.* 2004b). However effective pollination appears to be predominantly facilitated by birds (such as the swift parrot *Lathamus discolor,* the green rosella, *Platycercus caledonicus,* the yellow wattle bird *Anthochaera chrysoptera,* the yellow-throated honeyeater *Lichenostomus flavicollus,* the black-headed honeyeater *Melithriptus affinus,* and the new holland honeyeater *Phylidonryris novaehollandiae* (Hingston and Potts 1998a; Hingston *et al.* 2004a; Hingston *et al.* 2004b). The study site at the Tinderbox Hills was the nesting site of a pair of green rosellas, *Platycercus caledonicus,* for three consecutive seasons over the course of the study, which are active pollinators of *E. globulus* (Hingston *et al.* 2004a).

The leptokurtic decrease in pollination success associated with distance correlates well with the observed avian foraging pattern displayed by pollinators of *E. stoatei* (Hopper and Moran 1981) and *E. unigera* (Savva *et al.* 1988; Potts 1990)). The feeding behavior of pollen and nectar feeding parrots discussed in Southerton *et al.* (2004) suggest that nearest neighbor foraging movements are interspersed with much longer distance movement to other feeding areas, hence allowing for pollen dispersal over tens of kilometers. Under this model, the likelihood of long distance pollen flow (less than one kilometer) into the present study site does exist, but is minimal in comparison to short distance (< 15 m) dispersal. This would be suggested by extrapolation of the pollen dispersal curve of this population sample, with an expected very long tail pollen dispersal distribution similar to that displayed by *E. nitens* which is predominantly insect

pollinated (Barbour *et al.* 2005). In the intensive study of interspecific hybrids, the number of pollination events from the *E. nitens* source dropped rapidly to 0.7% of the individuals sampled within the 200 to 300 m distance class, with a similar low percentage (rare pollen dispersal events) being maintained at all sampling distances up to 1.6 km away from the source population. Pollen dispersal distances have been documented in natural populations of a number of eucalypt species including *E. rhodantha* (maximum < 170 m (Sampson *et al.* 1989)), *E. pulverata* (maximum = 250 m (Pryor 1976; Peters *et al.* 1990)), *E. risdonii* (maximum > 500 m (Potts and Reid. 1988)) and *E. regnans*, (maximum = six km (Ashton and Sandiford 1988)). However, comparisons of pollen flow distances between species must be made with caution, as estimates of average and maximum pollen distances are directly related to the scale of the individual study and numerous other factors.

There was no significant directional effect between potential parents on the likelihood of pollination success (using the Neighborhood model with the software of Burczyk 2002). Likewise, there was no significant effect of the interaction between distance and angle on the likelihood of pollination success. However, when the angle between assigned parent couples and their naturally established offspring in the juvenile cohort at the study site was investigated, a clear pattern emerged. Seedlings were predominantly located to the east of their assigned parents. In this case, the father and mother are included in the comparison, and so this angle is the combined effect of both pollen and seed dispersal. Infact, the predominant direction between assigned parents and established offspring in this analysis was similar to that expressed by the pattern of up slope allelic movement revealed with the cohort wide comparisons made in Chapter 2. In both analyses, either an east or north-easterly genetic shift from mature to juvenile cohorts is displayed. This is the direction of the predominant westerly airstream in Tasmania (Jackson 1981) which is particularly prevalent at the study site. There was no predominant angle of pollen movement revealed by the paternity analysis of the nine open pollinated families. It is much more likely that seed dispersal as opposed to the movement of bird and insect pollinators (Hingston and Potts 1998a) is influenced by the predominant west to southwesterly winds that this population experiences. The seed of *E. globulus* displays no specialized adaptation to dispersal and so is only affected by wind and gravity

(Cremer 1977; Potts and Wiltshire 1997) and is dispersed either as the seed capsule (woody fruit) dries out, dehisces and releases seed, the entire capsule detaches from the tree, or a branch (with capsules attached) detaches from the tree. Numerous field observations at the Tinderbox study site (T. Jones) suggest that a much higher rate of capsule dispersal seems to occur under windy conditions (in the absence of fire), as branches and capsules are broken off and blown away from the tree. Indeed, Potts and Wiltshire (1997) have observed clear asymmetry in seed dispersal in *E. risdonii* in Tasmania, with seed movement following the direction of the prevailing westerly wind. It is therefore more likely that the up slope geographic shift in allele frequencies between the mature and juvenile cohort is likely due to directional seed movement (assuming similar conditions when the seedlings of the juvenile cohort were established). This situation contrasts with that observed in wind pollinated species that often display directional pollen movement, such as *Pseudotsuga menziesii* (Burczyk and Prat 1997), *Pinus attenuata* (Burczyk *et al.* 1996), *Quercus macrocarpa* (Dow 1995, cited in Dow and Ashley (1998), *Quercus robur* and *Quercus petraea* (Streiff *et al.* 1999), *Pinus densiflora* (Lian *et al.* 2001).

Comparison of the pollen dispersal curve with the total gene dispersal curve generated from the assignment of parents in the mature cohort to seedlings in the juvenile cohort can provide insights into the shape of the seed dispersal curve alone. As it is not possible to differentiate between mother and father in parentage assignment, without maternal markers, the position between seedling and assigned parent can be considered a function of either seed dispersal alone in the case of maternal gene flow, or pollen dispersal to the mother and then seed dispersal in the case of paternal gene flow. The combined gene dispersal curve displayed a much higher frequency of dispersal events between 20 and 40 m than the pollen dispersal curve indicating that realized seed dispersal may be occurring frequently at these distances, relative to pollen movement over the same distance. This suggests that regeneration niches may not be available closer to the parental trees. The effect of seed dispersal appears to have a broadening effect on the gene dispersal within the local site. However, while the decline in seed dispersal with distance may be less extreme than that expressed by the pollen dispersal curve, the potential for longer distance seed dispersal events (beyond the boundary of

the study site) is less likely due to the mechanisms of dispersal (Potts and Wiltshire 1997). Conceptually, the component of the total gene dispersal curve due to pollen flow could be calculated from the pollen dispersal curve and removed, leaving the effect of gene dispersal due to seed alone. However, the noise within the total gene dispersal curve resulted in a poor fitting regression curve, and this was not possible to undertake with a reasonable level of confidence. In addition, a number of factors must be taken into account when drawing conclusions from the comparison of the total gene dispersal curve and the pollen dispersal curve. Firstly, the mother trees sampled to create the pollen dispersal curve were very close to the centre of the site, whereas the juvenile seedlings that were assigned parents are distributed relatively evenly throughout the sample site. Secondly, the total gene flow curve is based on seedlings that have established in the population, whereas the pollen dispersal curve is based on seed that has not undergone field selection. Despite the pattern of seed dispersal in the population, the germination and establishment of *E. globulus* seedlings is dependent on the availability of resources in the place to where the seed is dispersed (Potts and Wiltshire 1997). Competition closer to the source (such as light and nutrient availability) may prevent establishment. Given the high density of mature trees, potentially few opportunities exist for successful establishment of seedlings and this is likely to play a major factor in the resulting pattern of seed dispersal determined with this approach. Pollen dispersal on the other hand, was determined by collecting seed from specific mother trees, and with the exception of pollen competition within the fruit (Pound *et al.* 2003), is not biased by such factors.

While the likelihood of pollination increases as the distance between mother and potential father decreases, the actual number of potential pollen donors at a particular distance decreases with proximity, in a population of even density (Crawford 1984). Consequently, pollen donors within close proximity to a mother tree are expected to pollinate with higher frequency than those further away (Wright 1943). This phenomenon was clearly displayed in the current study population where fathers closest to the mother generally sired more progeny within a family than those further away. This was illustrated by significant exponential decrease in the number of progeny sired from a particular father relative to the distance from the mother tree. The consequence

of this in terms of spatial genetic structure is an increased likelihood of full-sibling seedlings establishing in close proximity to each other.

The correlated paternity model of Ritland (2002) suggested that the likelihood that two randomly chosen progeny from an open pollinated family in this population (excluding self pollinated progeny) are full-siblings was 0.053 ± 0.008. This results in an estimate of effective population size (number of males contributing pollen to an open pollinated family) of $N_{ep}$ = 19.23 pollen donors. However using Ritland's (1989) correlation of paternity ($r_p$) to estimate effective population size is prone to under-estimation (Hardy *et al.* 2004) as is suggested by the Neighborhood approach of Burzyck (2002) which estimated $N_{ep}$ = 23.39 (19.79% of potential parents within the mature cohort) in the current study. The Neighborhood method estimate is not taking into account individuals from outside the study site nor potential father genotypes that were excluded due to the inability of Neighbor 2.0 to deal with missing data, and so it is likely that this approach is also underestimating the effective population size. Estimates of effective population size in open pollinate seed of other native forest species populations range from 2.9 in *Albizia julibrissin* (Irwin *et al.* 2003) to 125 in *Pinus sylvestrus* (Robledo-Arnuncio *et al.* 2004) with effective population size for a further eight native populations of forest species listed in Hardy *et al.* (2004)) predominantly below 10. Sampson (1998) reports an effective population size in *E. rameliana* of 11. Once again, estimates of effective population size are dependent on the sample scheme of the study such as quantity and position of seed sampling within the canopy (Patterson *et al.* (2001; 2004), so comparison between studies should be made with caution. Patterson *et al.* (2004) showed that outcrossing is much higher in the upper canopy than the lower branches and, as suggested by Patterson *et al.* (2004), seed was collected from the upper canopy in the current study to maximum outcrossing potential.

When considering the significantly higher relatedness of individuals within close proximity to each other revealed in Chapter 2, and the high likelihood of near neighbor pollination events displayed by the pollen dispersal curve of the current study, it is likely that a significant proportion of mating events may occur between related individuals in this population, as suggested by numerous studies including Gliddon *et al.* (1987), Hamrick and Nason (1996) and Hedrick (1990). A significant proportion of the open

pollinated progeny genotyped in this study displayed some level of bi-parental inbreeding. At least 13% of the out-crossed progeny where assigned to parents that shared relatedness at the level of half-siblings. One percent of the assigned out-crossed progeny were the offspring of parents that displayed relatedness values equivalent to full-siblings.

The frequency distribution of pair-wise relatedness values of the assigned parents of the open pollinated families displayed a disproportionate peak in frequency at the value equivalent to half-siblings. There are two obvious possibilities as to why this is occurring. Firstly, from a proximity point of view, the likelihood of matings is greatest between trees that are very close to each other and trees in close proximity have a greater likelihood of being closely related (Chapter 2) . Secondly, flowering time in *E. globulus* is under strong genetic control (Gore and Potts 1995) and thus related trees are more likely to flower at the same time, again favoring mating between relatives. The lack of relatedness values of mating pairs equivalent to full-siblings most likely reflects the deficit of relatedness values equivalent to full-siblings in the mature cohort (Figure 9). Bi-parental inbreeding within the grand-parental population (the parents of the mature cohort that were assigned as parents of the open pollinated families) can be assessed by examining the pair-wise fraternity (four gene relatedness) between the assigned parents of the open pollinated families (Lynch and Ritland 1999). The tail of the frequency distribution of pair wise fraternity values between the assigned parents was slightly skewed towards positive values, indicating some level of bi-parental in-breeding in the parental population, however more exact quantification of this phenomenon was not achieved.

Significant variation in outcrossing rate between the nine families was evident in the current study, ranging from 0.61 to 1.00. Outcrossing rate has been previously shown to vary between *E. globulus* trees both in the native environment (0.48 to 1.00 Hardner *et al.* 1996 and the seed orchard situation (0.40 to 0.85, Patterson *et al.* 2004; 0.77, Moncur 1995; 0.79 to 0.92, Russell *et al.* 2001). The average out-crossing rate displayed in this population (0.90 ± 0.04) was the same as that found in a native dwarf population of *E. globulus* situated on the east coast of Tasmania (0.9, Foster *et al.* unpublished). The inter-tree variation in out-crossing rate amongst trees growing in close proximity (0.61-

1.00) was high as has been shown in other eucalypt populations (native and seed orchards), such as *E. regnans* (Griffin *et al.* 1987; Griffin and Cotterill 1988), *E. obliqua* (Brown *et al.* 1975), *E. leucoxylon* (Ellis and Sedgley 1993), *E. pulverulenta* (Peters *et al.* 1990), *E. grandis* (Campinhos *et al.* 1998) and *E. nitens* (Grosser *et al.* 2001).

Mating is rarely panmictic in eucalypt populations (Pryor 1976; Moran and Bell 1983; Eldridge *et al.* 1993) with consistent reports of heterozygote deficiency in seed cohorts in the genus. Potts and Wiltshire (1997) summarize the difference in observed heterozygosity and inbreeding coefficient ($F_{is}$) between seed and mature cohorts of six eucalypt species in the natural environment and two in the seed orchard situation. In almost all cases, the seed cohort displayed reduced observed heterozygosity and a higher inbreeding coefficient in comparison to the mature cohort. Potts and Wiltshire (1997) suggest that this deficit in heterozygosity could arise from either inbreeding and/or substructuring of populations (i.e. Wahlund effect). Despite the level of self pollination and bi-parental inbreeding described in the current study, surprisingly little variation in heterozygosity between the seed, juvenile, and mature cohorts was evident in the current study. There was a slight trend in the reduction of expected and observed heterozygosity between the seed cohort and later cohorts. Likewise, minor changes in the inbreeding coefficient between cohorts were displayed. A comparatively large number of individuals are contributing to the juvenile (at least 60 parents per 110 progeny) and seed cohorts (minimum average of 21.6 fathers per 60 progeny) in the study a population, and this appears to outweigh potential effects of self pollination and bi-parental inbreeding on the levels of heterozygosity in the population.

There was a decrease in allelic richness in the seed cohort, in comparison to the juvenile and mature cohort. This was most likely due to the limited number of mothers (nine) of the 549 individuals examined. It is interesting to note that the level of observed heterozygosity in the seed cohort was 0.80 (ranging from 0.64 to 0.87), despite 10% of the progeny being the product of self pollination.

The Australian *E. globulus* base breeding population was originally based on the assessment of open pollinated progeny collected from throughout the native distribution of *E. globulus* (Dutkowski *et al.* 1997). Mixed models incorporating best linear

unbiased predictions (BLUPS) are used to predict the breeding values of particular genotypes within this breeding population (Borralho and Dutkowski 1998) and selections within this breeding program have been made assuming an average level of selfing in base population open pollinated trees of 30% and that progeny are not bi-parentally inbred. If the local genetic dynamics quantified in the *E. globulus* population at the Tinderbox Hills are representative of the overall species, the level of self pollination assumed by the breeding program is quite inflated. Taking into account the findings of Tilyard *et al.* (unpublished) a self pollination rate of 30% will result in an average reduction of growth (stem basal area at six years) in an open pollinated family of around 11.4%. This is twice the level of inbreeding depression estimated to occur in the current study (5.6%) that also takes into account the effect of bi-parental inbreeding between half and full siblings. However, the current study site displays a very high density which is, in many cases, higher than that of the founding populations of the breeding program selections. It is likely that the selections originating from less dense populations display a higher level of inbreeding (Borralho and Potts 1996; Hardner *et al.* 1996). The potential for more serious bias in breeding values arises from the fact that levels of inbreeding may vary substantially between open pollinated families. In the case of growth, variable inbreeding depression will be confounded with breeding value estimates. In the present case, such variation is seen even between neighboring trees within a dense forest where outcrossing rates range from 60% to 100%. Similarly, these trees display significant variation in the levels of bi-parental inbreeding.

# Chapter 4: Genetic analysis of the Australian *Eucalyptus globulus* breeding population.

## Introduction

### Genetic diversity in domesticated species

The maintenance of a species potential for adaptation is regarded as an important direction in long-term conservation of tree species (Eriksson 2001; Hedrick 2004). Human impact on global forest genetic resources occurs via breeding and selection, seed relocation, population management and alteration of the environment (Lefèvre 2004). Tree breeding (and agricultural systems in general) and conservation of forest populations (and other plant populations) are not necessarily independent in their aims and outcomes (Eriksson *et al.* 1993; Bretting and Duvick 1997; Namkoong 1997; Yanchuk 2001). Both aim to maintain genetic diversity into the future to ensure long-term flexibility in response to changes in environmental and selective requirements. In addition to *in situ* native resource conservation, domestic biodiversity may provide a complimentary means of conserving genetic resources (Eriksson *et al.* 1993). Hence, correctly managed breeding programs of forest trees may be beneficial to maintaining genetic diversity and adaptive potential of forest species, in addition to promoting a successful breeding population (Yanchuk 2001; Lefèvre 2004).

In comparison to the domestication of many crop plants, beginning with the development of agricultural systems as many as ~10,000 years ago (Harlan 1992), the domestication of Australian native tree species is in its infancy (Moran *et al.* 2000; Potts *et al.* 2004). Consequently, cultivated crop systems can be used as a predictive model for other species in the early stages of domestication (Lefèvre 2004), providing potential directions for breeding strategies to achieve long term genetic diversity within the breeding populations, in combination with directed selection for breeding traits.

It is generally assumed that cultivated crop plants tend to display less genetic diversity than their corresponding wild type ancestors (Doebley 1989). During the course of

domestication, a species is usually spread to regions that display different ecological conditions to its native origin (Harlan 1992). The adaptation of a domesticated species to its new environmental conditions may select for genetic deviation from the wild type, in addition to the directional selective pressures based on the specific needs of man. However, divergent selection for optimal performance in different environments tends to result in populations of domesticated species displaying higher inter-population differentiation than wild populations (Doebley 1989). Furthermore, the combination of this phenomenon and admixture (crossing between cultivars) may maintain a high level of genetic diversity in breeding populations of domesticated species (e.g. pepper, Hernandez-Verdugo *et al.* 2001).

Homogeneity within cultivated species may predominantly reside in the genes (and their flanking regions) that code for traits of interest to man (e.g. maize, Gaut *et al.*, 2000 and Wang *et al.*, 1999) and, the actual genetic control of these traits may be due to very few actual genes (Paterson *et al.* 1995; Grandillo and Tanksley 1996; Doebley *et al.* 1997; Xiong *et al.* 1999; Poncet *et al.* 2000; Paterson 2002). In eggplant for example only six quantitative trait loci (QTL) are responsible for the major phenotypic differences between wild and cultivated varieties (with another 56 loci identified as contributing minor combined effects) (Doganlar *et al.* 2002). The remainder of the genome however, may be effectively neutral, maintaining a surprisingly high level of genetic diversity even in highly domesticated species (Wang *et al.* 1999; Gaut *et al.* 2000). In addition, genetic diversity within a domesticated species can be improved by careful management, as has been the case in wheat, with an increase in global genetic diversity occurring in the species between 1990 and 1997 (Reif *et al.* 2005), via the introgression of new alleles. In many cases, factors other than the specific process of domestication should be considered to explain the reduction in diversity seen within some crop plants (Lefèvre 2004), such as the original genetic diversity of the ancestral type (e.g. tomato, Miller and Tanksley 1990).

Nevertheless, the level of genetic diversity in tree breeding populations appears to be species dependent, and many breeding populations, particularly conifer species, display significantly lower genetic diversity than that of native populations (Szmidt and Muona 1985; Moran and Bell 1987; Williams and Hamrick 1995; Rajora 1999). Intensive

selection with the use of best linear unbiased prediction for example, generally results in a rapid short term response within a breeding program, but can cause a rapid reduction in effective population size and a decrease in the potential for long-term response (Caballero and Santiago 1994; Verrier *et al.* 1994). A reduction in effective population size will lead to a reduction in rare alleles (Nei *et al.* 1975). For example, allelic richness has been significantly reduced in a number of conifer seed orchards even when overall genetic diversity remains minimally affected (Godt *et al.* 2001). However, a larger breeding population, with the less intense selection required to maintain rare alleles, will take many more generations (10-20) to express improvement (Namkoong *et al.* 1988) and in tree species such as *E. globulus*, this means a long time, relative to other agricultural systems.

With the domestication of Australian tree species, including *E. globulus,* in its infancy, tree breeders are in a unique position to utilize current molecular marker technology to assess the genetic resources of their particular breeding programs (Moran *et al.* 2000). This has been carried out to some extent in a number of domesticated Australian species, indicating significant variability in the amount of genetic diversity within breeding populations and seed orchards. The domestication of the tropical species *Acacia mangium* (Butcher *et al.* 1998b), *A. aulacocarpa* (McGranahan *et al.* 1997), *A. auriculiformis* (Wickneswari and Norwati 1993) and *Eucalyptus pellita* (House and Bell 1996) appears to have resulted in the inclusion of an estimated 56%, ~36%, 80% and 67% respectively of the native genetic diversity into breeding populations as reviewed in (Moran *et al.* 2000). However, comparisons of genetic diversity between populations must be carried out with caution, ensuring that assays use the same molecular markers for both populations and take into account sample size to avoid bias (Lefèvre 2004). While assessment of genetic variation within *Eucalyptus* has been carried out in a number of species, both in native populations (e.g. *E. cladocalyx*, McDonald *et al.* 2003, *E. nitens,* Byrne *et al.* 1998, *E. globulus*, Jones *et al.* 2002, *E. urophylla*, House and Bell 1994) and breeding populations (e.g. *E. urophylla*, Leite 1997 and Leite *et al.* 2002, *E. globulus*, Astorga *et al.* 2004 and Gemas *et al.* 2004, *E. grandis*, Leite 1997) very few comparisons have been made directly comparing breeding and native populations of the same species.

**Domestication of *Eucalyptus globulus*.**

*Eucalyptus globulus* was initially documented by French explorers in 1792 (Potts *et al.* 2004) and formerly described in 1799 (Labillardière 1799), before being introduced to Europe (France) in 1804. By the end of the 19th century, the species had spread to Chile, South Africa, Portugal, Italy, India, Spain, USA, Uruguay, Algeria, Tunisia, Argentina, Peru, Ecuador, Zimbabwe, China, Ethiopia and Bolivia (Potts *et al.* 2004) and the process of domestication had begun. The first formal breeding program for the species began in 1966 in Portugal and by the late 1985, more than 20 trials, based on an extensive sampling of the native distribution of the species (Orme 1977) from Tasmania and southeastern mainland Australia, had been established in Australia, and overseas (Potts *et al.* 2004). In 1987, another collection was undertaken by the CSIRO Australian Tree Seed Centre, collecting seed from 616 trees from 49 localities throughout the native distribution of the species and trials based on this material now exist in Australia (Jarvis and Borralho 1995), Chile (Infante and Prado 1989; Infante and Prado 1991; Ipinza *et al.* 1994; Sanhueza and Griffin 2001), China (Zang *et al.* 1995), Ethiopia (Gizachew 2002), Portugal (Araújo *et al.* 1996) and Spain (Vega Alonso *et al.* 1994; Soria *et al.* 1998; Toro *et al.* 1998). Breeding programs for the species, some of which are now into their third generation, are present in Australia, Ethiopia, Argentina, Uruguay, Spain, India, Portugal, and Chile (Potts *et al.* 2004). The global plantation area of *E. globulus* may now have reached 2.5 million hectares (Potts *et al.* 2004).

Optimal domestication of the species will involve the directional selection from the genetic resource (Harwood 1999) without compromising the ability to adapt to a changing environment (Moran *et al.* 2000; Burley 2001). Investigations into genetic diversity within two Portuguese breeding populations of *E. globulus*, based on Portuguese land races (established local *E. globulus* populations in Portugal) and Australian native stand collections (Gemas *et al.* 2004) have been carried out, using inter sample sequence-repeat (ISSRs). This study identified significant genetic variation within both individual land races and breeding population races, between land races and between breeding population races. However, with no direct comparison to the wider *in situ* native distribution of *E. globulus*, comparisons in this case have been limited to between breeding program selections and samples of open pollinated seed from the

original localities that comprise base population samples. The native distribution of *E. globulus* has been extensively sampled by Jones *et al.* (2002) and Steane *et al.* (in prep) using microsatellite markers. This has provided a benchmark of genetic diversity within the species that is suitable for direct comparison with breeding population genetic diversity if the same markers are used.

This study aimed to evaluate the genetic diversity within the Australian *E. globulus* breeding population in comparison to that seen in the wild. Microsatellite loci were used to assess genotypes within the breeding population which comprises seven out of the thirteen native races classified by Dutkowski *et al.* (1999). The diversity within this sample was compared to that found in a large sample of mature trees in the native distribution of *E. globulus* in Tasmanian and southeastern mainland Australia (Steane *et al.* in prep) that included individuals from 10 of the thirteen native races of Dutkowski *et al.* (1999). Three small atypical races (Lighthouse, Dromedary and Recherche Bay) were excluded. In addition, genetic assignment (Pritchard *et al.* 2000a) and parentage analysis (Ritland 2002) were used to validate the classification of race and family of individuals within the breeding population, to gain an overall estimation of the level of pedigree errors occurring within the breeding population.

## Materials and Methods

### Plant material and overview

The material used in this study consisted of a sample of 149 individual genotypes, comprising first generation selections from base population trials of the Australian *E. globulus* base breeding population, managed by the Southern Tree Breeding Association. The sample consisted of material originating from seven of the native *E. globulus* races classified by Dutkowski *et al.* (1999), with the genotypes originating from 80 open pollinated families from 26 native stand localities. Material was sampled from seed orchards and arboreta situated in Northern Tasmania, Southeastern Australia and Western Australia. DNA was extracted from leaf tissue as described in Chapter 2. Eight microsatellite markers were used in the analysis of the breeding population sample, seven of which corresponded to those used by Steane *et al.* (in prep) to genotype the sample of the native *E. globulus* distribution (340 mature strata individuals from 10 races and 46 localities). The eighth locus was included to allow even more detailed analysis within the breeding population sample.

Firstly, the genetic variability within the breeding population was compared to that found in the native distribution of *E. globulus* (Steane *et al.* in prep). In addition, a nested analysis of molecular variance (AMOVA) was used to partition the genotypic variation between race, locality within race, family within locality and individual within family. Secondly, the breeding program classification of race and family of each individual sampled was verified where possible, using Bayesian assignment tests (Pritchard *et al.* 2000a; Rosenberg *et al.* 2003) to probabilistically verify race, and parentage analysis (Ritland 2002; Jones and Ardren 2003) where multiple samples had been made from within a family. An attempt to define erroneous individuals within a single family classification was also made using a clustering approach with principal coordinate analysis (Peakall and Smouse 2001).

In addition to the 149 genotypes examined, further clonal replicates of six genotypes from the breeding population were included in the analysis as these individual ramets displayed some phenological differences (e.g. flowering time) that were not expected

between clones of the same genotype. Clonal replicates were not included in any of the analyses, except when their genotypes were compared to check for classification error.

**Microsatellite markers**

The eight microsatellite loci used in this study (EMCRC2, EMCRC3, EMCRC6, EMCRC7, EMCRC10, EMCRC11, EMCRC12, EMCRC5) were developed in *E. globulus* by Steane *et al.* (2000). EMCRC5 was only used to genotype the breeding population sample and so analyses involving only the breeding population were carried out with the full eight loci, in comparison to the seven loci used in comparing both breeding and native population samples. Hardy Weinberg equilibrium for each locus was tested using the combined dataset of the native distribution of *E. globulus* and the breeding population sample, and the breeding population sample only, to test EMCRC5. Hardy Weinberg equilibrium was tested within and overall all races, with the corresponding races of the native and breeding population samples treated as independent samples. Tests were randomization based and carried out with the software Fstat V. 2.9.3.2 (Goudet 1995). Alleles within race were used as the unit of re-sampling, testing the statistic $F_{it}$ (Goudet 1995; Weir and Hill 2002). Tests for linkage disequilibrium between loci were also carried out with Fstat V. 2.9.3.2, based on 560 permutations using the G statistic (Goudet *et al.* 1996). While being very sensitive, the G statistic has an advantage over Fisher's exact test as each sample is weighted by its information content (Goudet *et al.* 1996).

**Genetic comparison between native and breeding population samples**

The overall genetic diversity in the breeding population sample was initially compared to that within the native distribution of *E. globulus*, firstly using the same races as those in the breeding population sample, and secondly, using the total native distribution of *E. globulus* sampled so far (Steane *et al.* unpublished). $F_{st}$ (Weir and Cockerham 1984) between the breeding population sample and the native distribution samples was calculated, and the significance of genetic differentiation between the two samples were

calculated using the maximum likelihood G statistic (over 10000 permutations with Fstat V. 2.9.3.2). Observed heterozygosity, expected heterozygosity, private alleles (per sample, standardized for sample sizes by dividing by the number of individuals in the sample), allelic richness (alleles per locus, standardized to the number of individuals within smallest sample El Mousadik 1996) and Wright's inbreeding coefficient, $F_{is}$ (Weir and Cockerham 1984), were calculated for the three samples, using the seven loci in common with all samples, using Fstat V. 2.9.3.2 and GDA V. 1.2 (Lewis and Zaykin 2001).

A hierarchical comparison of genetic structure within the native distribution samples and the breeding population sample was also made, using Wright's F statistics (Weir and Cockerham 1984) to summarize the amount of genetic variation between individuals within race ($F_{is}$), between individuals within the total sample, ($F_{it}$) and the genetic variation between races within each sample ($F_{st}$) using Fstat V. 2.9.3.2 and GDA V.1.2.

Similar descriptive analyses were carried out between the seven races within the breeding population. In addition, pair-wise $F_{st}$ (Weir and Cockerham 1984) and the significance of genetic differentiation using the maximum likelihood G statistic between the seven races within the breeding population and their corresponding races in the native distribution sample was calculated with the software Fstat V. 2.9.3.2. Twenty thousand randomizations were used, permuting genotypes among samples, assuming significant deviation from Hardy Weinberg equilibrium.

The comparison between native and breeding population racial variation was summarized using UPGMA (Lewis and Zaykin 2001) clustering to produce a tree of stepwise genetic differentiation between races based on pair-wise $F_{st}$ (coancestry coefficient). This was carried out using GDA V. 1.1.

**Genetic differentiation within the breeding population**

Genetic differentiation between the seven races within the breeding population sample was investigated, using both Nei's genetic distance (1978) and pair-wise $F_{st}$ between groups with GDA V. 1.1. Genetic differentiation between the seven races was also tested with the G statistic. Twenty thousand randomizations were again used, permuting genotypes among races, assuming significant deviation from Hardy Weinberg equilibrium. Again, UPGMA clustering was used to produce a tree of hierarchical, stepwise genetic differentiation between the seven races within the breeding population sample, based on pair-wise $F_{st}$ (coancestry coefficient). This was carried out using GDA V. 1.1.

**Analysis of molecular variance**

Analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992; Michalakis and Excoffier 1996) was used to illustrate the partitioning of genetic variation between the race, locality and family classifications of the breeding population. Two three level AMOVA analyses were carried out, using GenAlEx V 5.1 (Peakall and Smouse 2001). The first analysis quantified the proportion of genetic variation in the sample due to differences between races, localities within races and between individuals within localities. The second analysis examined genetic variation between localities, families within localities and individuals within families. The results of the two analyses were combined by partitioning the variance between individuals within locality of the first analysis, into that due to families within localities and individuals within families, gained from the second analysis. Only families that consisted of two or more individuals were included in this analysis, resulting in a sample size of 90 individuals, with individuals within families ranging from two to six, with no clonal replication.

**Checking for pedigree errors**

The first step of the error checking process was to verify the racial classification of each individual within the breeding population. This was initially carried out using a Bayesian assignment approach with the software Structure 2.0 (Pritchard *et al.* 2000a) (see Chapter 2 for a detailed description of this approach). Gene frequencies for each race were calculated from the predefined racial classifications of the breeding population, followed by the probability based re-assignment of individuals to the seven races. A burnin period of 10000 repetitions was used, followed by 10000 Monte Carlo repetitions, assuming allele frequency correlation among populations and different $F_{st}$ values within different races.

A similar approach to verify locality classification was not possible, due to the low number of individuals within the large number of localities. However, verification that individuals of the same classified family within the breeding population sample actually share at least one parent was possible. Sixteen families that included between three and six individuals were analyzed for a common parent using MLTR V. 3.0 (Ritland 2002).

With such low numbers of repetition within family, it is not immediately clear as to which individuals are erroneously classified in this case as it is difficult to define the common parental genotype. Genetic variation between individuals within the erroneous families was then examined using principal coordinate analysis (PCO Sokal and Rohlf, 1981) carried out with GenAlEx V5 (Peakall *et al.* 1995; Peakall and Smouse 2001). In this case, individuals that share the same parent will occur in close proximity to each other on one or more principal coordinate axes as they all share a close genetic distance to one another (Sokal and Rohlf 1981), and those not part of the family are likely to be more distant from the family cluster. If there is, for example, one erroneous member of the classified family, it should be situated further away from the real family cluster. All individuals within the breeding population were used in the ordination.

Between two and four clonal replicates of six individual genotypes (1 x 4 repeats, 2 x 3 repeats, 3 x 2 repeats) included in the breeding population sample were compared for error in genotype classification, by visually analyzing alleles at the eight loci for mismatches.

# Results

## Linkage and Hardy Weinberg equilibrium

Highly significant deviation from Hardy Weinberg equilibrium was detected in six out of eight loci after the Bonferroni correction (Rice 1989) for multiple comparisons in the within race analysis (Table 4.1). All loci displayed significant deviation ($P < 0.0001$) from Hardy Weinberg equilibrium when race was ignored. No significant genetic linkage between any pairs of loci was detected after the Bonferroni adjustment (Rice 1989), however, the probability was very close to the 0.05 threshold in a number of cases.

Table 4.1: Number of individuals (*N ind*) genotyped with the particular locus in the combined dataset of all races within the breeding population and the 10 native population races, number of alleles (*N alleles*), expected heterozygosity ($H_e$), observed heterozygosity ($H_o$), Wrights F statistics ($F_{is}$, $F_{it}$, $F_{st}$) and probability levels for the test of deviation from Hardy Weinberg equilibrium (within race analysis) for the 8 loci used in the study. The probability threshold for all loci excluding EMCRC 5 at the 0.05% level is equal to 0.007 after the Bonferroni adjustment for multiple comparisons. ¥The corresponding probability threshold for EMCRC5 is equal to 0.006 as this locus was analyzed in the 8 locus analysis of the breeding population sample only. Other statistics for EMCRC5 are based on the breeding population sample alone. These parameters were drawn from the pooled natural and breeding population data

| Locus | *N ind* | *N alleles* | $H_e$ | $H_o$ | $F_{is}$ | $F_{it}$ | $F_{st}$ | *HW Prob* |
|---|---|---|---|---|---|---|---|---|
| EMCRC2 | 465 | 22 | 0.83 | 0.68 | 0.10 | 0.19 | 0.10 | 0.000 |
| EMCRC3 | 474 | 15 | 0.74 | 0.57 | 0.12 | 0.23 | 0.13 | 0.000 |
| EMCRC6 | 459 | 22 | 0.83 | 0.71 | 0.10 | 0.14 | 0.05 | 0.000 |
| EMCRC7 | 458 | 20 | 0.77 | 0.66 | 0.09 | 0.16 | 0.07 | 0.001 |
| EMCRC10 | 455 | 17 | 0.85 | 0.59 | 0.25 | 0.30 | 0.07 | 0.000 |
| EMCRC11 | 476 | 21 | 0.89 | 0.80 | 0.03 | 0.10 | 0.08 | 0.100 |
| EMCRC12 | 475 | 20 | 0.77 | 0.68 | 0.05 | 0.12 | 0.08 | 0.056 |
| EMCRC5¥ | 145 | 34 | 0.82 | 0.55 | 0.28 | 0.32 | 0.08 | 0.000 |
| Combined | 466 | 21.4 | 0.81 | 0.67 | 0.13 | 0.17 | 0.08 | 0.000 |

## Genetic comparison between native and breeding population samples

Comparison of the native distribution samples (one sample including only the seven races corresponding to those in the breeding population sample and the other including 10 native races) with the STBA breeding population sample resulted in a very low global $F_{st}$ between the natural stand and breeding population samples (0.008 and 0.009 respectively). Both analyses revealed highly significant genetic differentiation between

the native and breeding population samples ($p > 0.000$) using the G statistic of Goudet (1996). This indicates the sensitivity of the G statistic in this case and suggests that caution must be taken when using this statistic to test for population differentiation.

The breeding population sample of seven races displayed higher observed genetic heterozygosity ($H_o$) than both the independent sample of the seven races based on native stand trees (n = 231) and the most complete sample of the native distribution of *E. globulus* to date (10 races, n = 340) (Table 4.2). Interestingly however, there is little difference in expected heterozygosity between samples, and a lower allelic richness and proportion of private alleles (per sample) in the breeding population sample. The breeding population sample exhibited a lower $F_{is}$ than the native distribution samples, indicating a lower level of 'inbreeding' within races of the breeding population sample, although this could also be due to geographic structuring. However, less variation in $F_{it}$ and $F_{st}$ was seen between the breeding population sample and the two native population samples.

Table 4.2. Comparison of genetic diversity indices between the breeding population sample and the two native distribution samples. ***N*** = number of individuals within sample, ***He*** = expected heterozygosity, ***Ho*** = observed heterozygosity, ***Pa*** = private alleles (per sample, but divided by the number of individuals to account for variation in sample size), $\mu$ = mean number of alleles per locus, $R_t$ = allelic richness (standardized to the size of the breeding population), $F_{it}$, $F_{is}$, $F_{st}$ = Wrights F statistics (Weir and Cockerham 1984).

| Sample | *N* | *He* | *Ho* | *Pa* | $\mu$ | $R_t$ | $F_{is}$ | $F_{it}$ | $F_{st}$ |
|---|---|---|---|---|---|---|---|---|---|
| Native (7 races) | 231 | 0.80 | 0.66 | 0.11 | 17.14 | 16.31 | 0.118 | 0.194 | 0.086 |
| Native (10 races) | 340 | 0.81 | 0.66 | 0.11 | 18.71 | 16.72 | 0.128 | 0.196 | 0.078 |
| Breeding | 149 | 0.82 | 0.71 | 0.04 | 14.28 | 14.26 | 0.059 | 0.145 | 0.092 |

Each race within the breeding population sample was genetically very similar to that of the corresponding race within the native distribution sample. In the hierarchical UPGMA analysis, each race from the breeding population sample clustered closely with its corresponding race within the native distribution sample (Figure 4.1), with the exception of the Southeastern Tasmanian race. Further genetic affinities between the various races have been closely examined by Steane et al. (in prep).

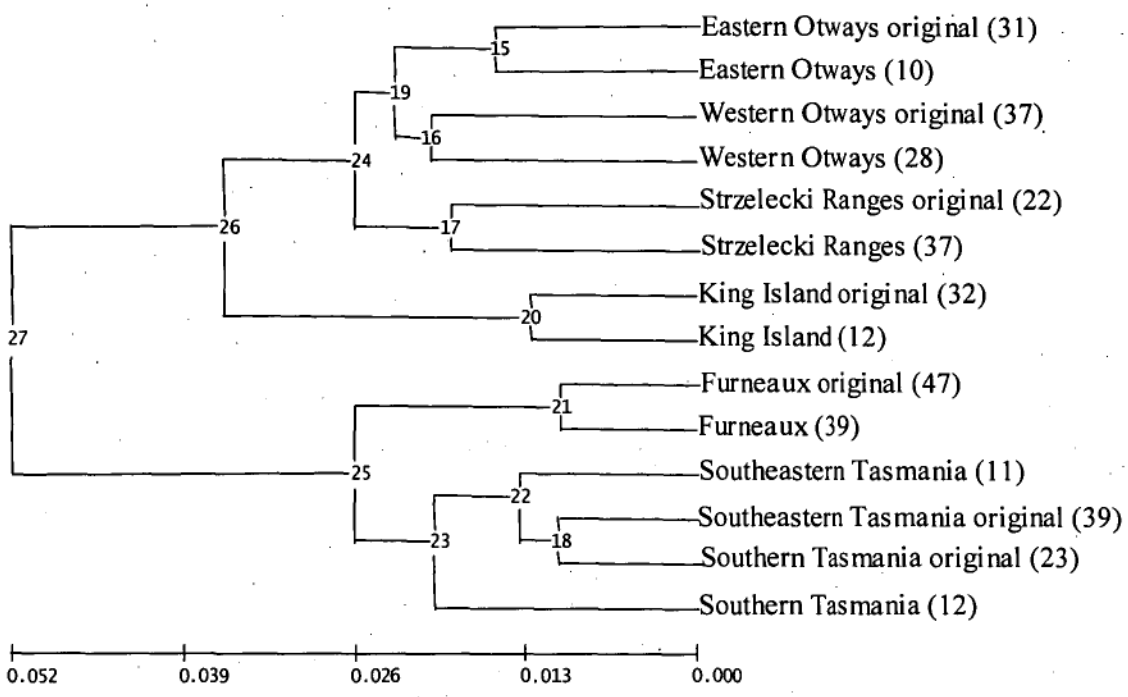



Figure 4.1: The UPGMA clustering dendrogram displays the hierarchical genetic affinities, based on pair-wise $F_{st}$ (coancestry coefficient) (Weir and Cockerham 1984), between the seven races within the breeding population sample and the corresponding races within the native (original) distribution sample. The number of individuals in each race sample are provided in brackets.

The genetic differentiation explained by pair-wise $F_{st}$ between the corresponding races of the breeding population and native distribution samples was very small in each case, with most values very close to zero (Table 4.3). Despite low pair-wise $F_{st}$ values, both the Furneaux and Southeastern Tasmanian races displayed significant genetic variation between breeding population and native distribution samples, based on Goudet's maximum likelihood G statistic. All races except the Southern Tasmanian race within the breeding population sample expressed slightly higher (around five meters) observed genetic heterozygosity than the corresponding races from the native distribution sample (Table 4.3), however this trend was not evident when examining expected heterozygosity between samples (Table 4.3).

**Genetic differentiation within the breeding population**

While the average number of alleles per locus ranged from 5.9 to 11.3 between the races within the breeding population, allelic richness only ranged from 5.3 to 6.7 when standardizing mean allele number to the race of the lowest number of individuals (Eastern Otways, $N = 10$) (Table 4.3). The King Island race displayed the lowest genetic diversity of the seven races within the breeding population, displaying the lowest allelic richness, expected and observed heterozygosity, and a low $F_{is}$ (Table 4.3). However, individuals within the race are only represented by one locality in this sample, consistent with the native stand sample. The other races displayed various ratios of these diversity indices, with no race clearly appearing the most genetically diverse (Table 4.3).

Table 4.3. Comparison of genetic diversity indices within races of the breeding population sample. $N_i$ = number of individuals within race, $N_l$ = number of localities represented within each race, $N_f$ = number of families within race, $\mu$ = average number of alleles per locus, = allelic richness (standardized to the smallest race sample), ***He*** = expected heterozygosity, ***HeND*** = expected heterozygosity in the native distribution sample, ***Ho*** = observed heterozygosity, ***HoND***, observed heterozygosity in the native distribution sample, $R_t$ = allelic richness (standardized to the size of the race), $F_{is}$ = heterozygote deficit of individuals within races of the breeding population (Wrights inbreeding coefficient). The differentiation between the breeding population sample and corresponding native population sample ($F_{st}$) of each race is shown, with the significance of this differentiation tested with Goudet's maximum likelihood G statistic (Goudet 1996). * = P < 0.05, *** = P < 0.001

| Race | $N_i$ | $N_l$ | $N_f$ | $\mu$ | $R_t$ | *He* | *He ND* | *Ho* | *HoND* | $F_{is}$ | $F_{st}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Eastern Otways | 10 | 3 | 8 | 6.6 | 6.4 | 0.77 | 0.74 | 0.70 | 0.66 | 0.101 | -0.002 |
| Western Otways | 28 | 3 | 15 | 8.9 | 6.3 | 0.77 | 0.75 | 0.71 | 0.68 | 0.076 | 0.011 |
| Strzelecki Ranges | 37 | 3 | 18 | 10.1 | 6.4 | 0.73 | 0.74 | 0.70 | 0.65 | 0.060 | 0.011 |
| Furneaux | 39 | 6 | 20 | 11.3 | 6.7 | 0.77 | 0.75 | 0.70 | 0.65 | 0.094 | 0.021* |
| King Island | 12 | 1 | 7 | 5.9 | 5.3 | 0.67 | 0.65 | 0.61 | 0.58 | 0.077 | 0.019 |
| Southeastern Tasmania | 11 | 4 | 5 | 6.1 | 5.8 | 0.80 | 0.79 | 0.67 | 0.66 | 0.158 | 0.054*** |
| Southern Tasmania | 12 | 4 | 7 | 6.9 | 6.3 | 0.80 | 0.81 | 0.72 | 0.74 | 0.103 | 0.014 |
| mean | 21.3 | 2.9 | 11.4 | 8.0 | 6.17 | 0.76 | 0.75 | 0.69 | 0.66 | 0.090 | 0.018 |

Almost all races were genetically differentiated from each other (Table 4.4), indicated by the matrix of pair-wise $F_{st}$ values and Nei's genetic identity (1978). Only two comparisons between races (Western Otways and Eastern Otways, Southeastern and Southern Tasmania) did not display significant genetic differentiation (using the G statistic), and in both these cases, the races being compared are in relatively close geographic proximity to each other and not separated by any geographic barrier.

Table 4.4: Nei's genetic distance (1978) with the significance of racial genetic differentiation, based on the G statistic, after the Bonferonni adjustment for multiple comparisons (Rice 1989) (above diagonal) and pair-wise $F_{st}$ (co-ancestry coefficient) (Weir and Cockerham 1984) (below diagonal) for comparisons between the seven races within the breeding population sample. * = P < 0.05, ** = P<0.01 *** = P < 0.001.

| | Eastern Otways | Furneaux | King Island | Southeastern Tasmania | Southern Tasmania | Strzelecki Ranges | Western Otways |
|---|---|---|---|---|---|---|---|
| **Eastern Otways** | | 0.49*** | 0.13* | 0.51*** | 0.32*** | 0.14*** | 0.03 |
| **Furneaux** | 0.11 | | 0.47*** | 0.13*** | 0.26*** | 0.55*** | 0.58*** |
| **King Island** | 0.05 | 0.13 | | 0.39*** | 0.30*** | 0.32*** | 0.14** |
| **Southeastern Tasmania** | 0.10 | 0.03 | 0.12 | | 0.13 | 0.64*** | 0.54*** |
| **Southern Tasmania** | 0.07 | 0.06 | 0.09 | 0.02 | | 0.64*** | 0.27*** |
| **Strzelecki Ranges** | 0.04 | 0.13 | 0.11 | 0.14 | 0.14 | | 0.18*** |
| **Western Otways** | 0.01 | 0.12 | 0.05 | 0.11 | 0.06 | 0.05 | |

When the between race pair-wise $F_{st}$ matrix is represented in a hierarchical, step-wise manner, two major clusters are evident (Figure 4.2). The three mainland Australian races (Eastern Otways, Western Otways and Strzelecki ranges) cluster closely. The King Island race (in western Bass Strait) is an outlier to this group. The other major group comprises the two Tasmanian races (Southeastern Tasmania and Southern Tasmania) and the Furneaux race (eastern Bass Strait).

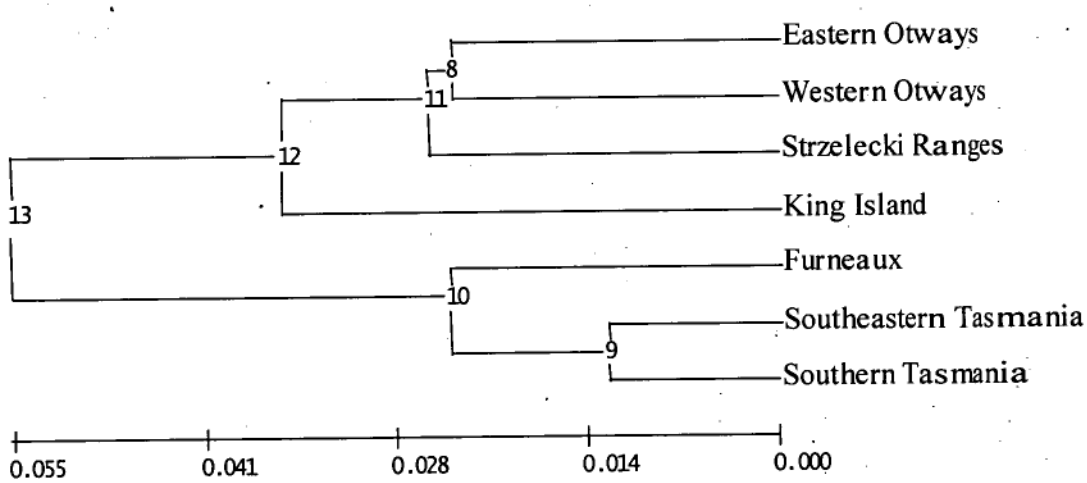



Figure 4.2: The UPGMA clustering dendrogram displays in a stepwise manner, the genetic differentiation between the seven races within the breeding population sample. Clustering was based on the pair-wise $F_{st}$ between individuals derived from eight microsatellite loci.

### Analysis of molecular variance

Racial differences accounted for 4.9% ($P < 0.01$) of the genetic variance within the breeding population sample ($N = 90$, excluding families with only one individual) (Figure 4.3). The difference between localities within races accounted for 5.8% ($P < 0.01$) of the genetic variance in the sample, with a further 8.7% ($P < 0.001$) of genetic variance accounted for by variation between families within localities (Figure 4.3). A remaining 80.6% ($P < 0.001$) of the genetic variance within the breeding population sample was accounted for by genetic variation between individuals within families (Figure 4.3).

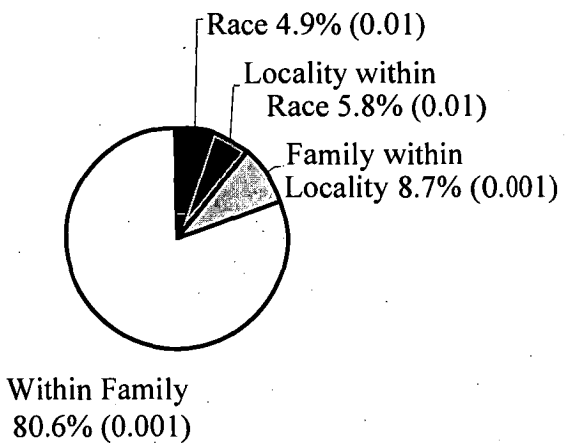



Figure 4.3: Analysis of molecular variance (AMOVA) of the breeding population sample. The AMOVA indicates the partitioning of genetic variance within the breeding population sample among races, localities, families and individuals. Families that were represented by only one individual were excluded from the analysis, leaving 90 individuals in total. The probability values for the significance of each effect are provided in brackets.

### Checking for pedigree errors

The probabilistic Bayesian based re-assignment of individuals to the seven races based on the gene frequency distribution within each race identified nine individuals within the breeding population that displayed a higher probability of belonging to a race other than that of their breeding population classification (Figure 4.4). In all but one instance, the difference between the most probable race, and the next most probable assignment was greater than 50% (Table 4.5), indicating a very large difference between the probability of the individual belonging to the breeding population classification and the newly

assigned race. It should be noted that a number of individuals within the Southern Tasmanian race classification have a noticeable probability of belonging to the Southeastern Tasmanian race, one of which displayed a higher than 50% probability of actually belonging to the Southeastern Tasmanian race. This is likely to be due to the close genetic affinity between the two races (Figure 4.2), rather than any error in classification. However, over 86% of individuals were independently assigned to their breeding population classified race with over 90% probability. In comparison, a similar analysis of the individuals sampled directly from the seven corresponding races in the wild displayed much less error (Figure 4.5), with only two individuals out of 231 displaying a marginally higher probability of coming from a different race to its real origin. One of these individuals was originally from Southeastern Tasmania and was assigned with 10% more probability to the Southern Tasmanian race than its real origin. The other was originally from the Western Otways race, and was assigned to the Eastern Otways race with only two percent greater probability. In both these situations, the misassigned race and the real race are in close geographic proximity to each other and are genetically similar (Table 4.4; Steane *et al.* in prep). This is a clear indication of the power of the Bayesian based assignment testing approach in this situation. Re-analysis of the genetic diversity indices and comparisons between races with the potentially erroneous individuals excluded resulted in slight reduction in expected and observed heterozygosities of the affected races, however the effect was minimal and did not justify presentation of the re-analyzed results.

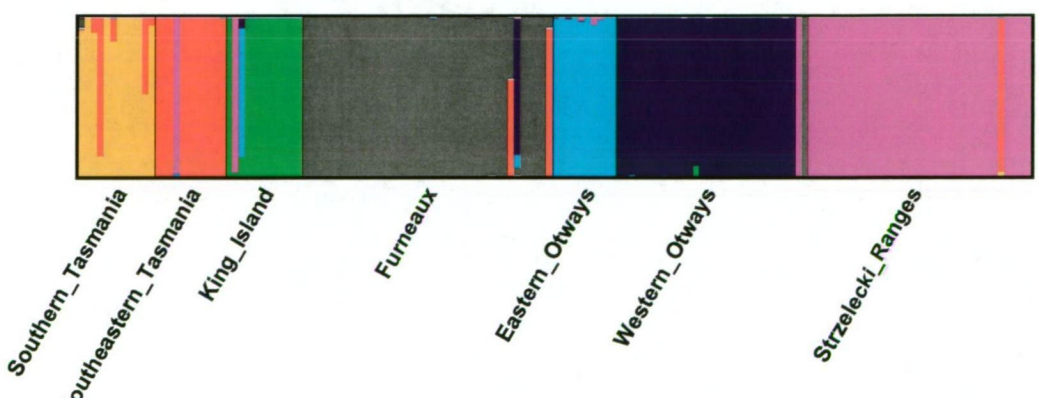



Figure 4.4: Assignment of individuals to the seven races of within the breeding population sample. Each individual ($N = 149$) is represented on the graph by a vertical line divided into coloured segments corresponding to the seven different races. The length of each coloured segment is proportional to the probability that an individual belongs to a race of the corresponding colour. Individuals are organized by their original racial classification (x axis).

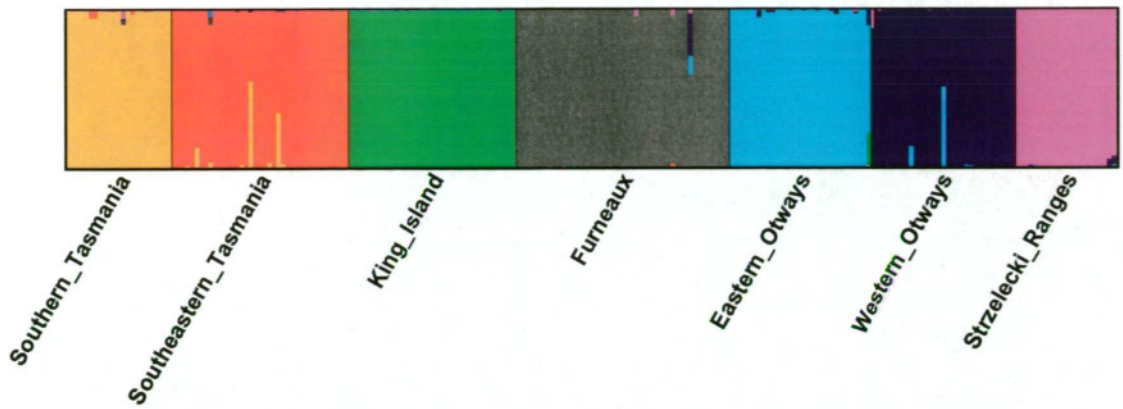



Figure 4.5: Assignment of individuals to the seven races within the native population sample that correspond to the breeding population sample. Each individual ($N = 231$) is represented on the graph by a vertical line divided into coloured segments corresponding to the seven different races. The length of each coloured segment is proportional to the probability that an individual belongs to a race of the corresponding colour. Individuals are organized by their original racial classification (x axis). Note that only two individuals displayed a marginally higher probability of originating from a different race to that of its real origin.

Table 4.4: The nine potentially erroneous individuals identified by the STRUCTURE analysis (Figure 4) are shown, with the corresponding probabilities of belonging to the seven breeding population races. The most likely race that these genotypes have originated from is indicated by the bold probability of assignment. Note that the probability of assignment for an individual does not add up to one over all races as the probability of assignment to a particular race is independently calculated based on the gene frequency distribution of the particular race (Pritchard *et al.* 2000a).

| STBA ID | STBA Race Classification | Eastern Otways | Western Otways | Strzelecki Ranges | Furneaux | King Island | Southeastern Tasmania | Southern Tasmania |
|---|---|---|---|---|---|---|---|---|
| 7991 | Furneaux | 0.00 | 0.00 | 0.00 | 0.33 | 0.00 | **0.51** | 0.01 |
| 8686 | Furneaux | 0.08 | **0.78** | 0.01 | 0.05 | 0.00 | 0.00 | 0.01 |
| 7902 | Furneaux | 0.00 | 0.00 | 0.00 | 0.06 | 0.00 | **0.77** | 0.00 |
| 12109 | King Island | 0.00 | 0.00 | **0.74** | 0.00 | 0.03 | 0.00 | 0.00 |
| 12110 | King Island | **0.67** | 0.05 | 0.00 | 0.00 | 0.11 | 0.00 | 0.00 |
| 5507 | S.E. Tasmania | 0.02 | 0.00 | **0.83** | 0.00 | 0.00 | 0.00 | 0.00 |
| 5797 | S. Tasmania | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.80** | 0.12 |
| 5598 | Strzelecki Ranges | 0.00 | 0.00 | 0.00 | **1.00** | 0.00 | 0.00 | 0.00 |
| 10188 | Strzelecki Ranges | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.96** | 0.03 |

Based on the eight loci used to genotype the breeding population sample, individuals of four out of the 16 classified families did not all share a common parent. While it was possible to determine families that contained some error using MLTR V. 3.0 (Ritland 2002), it was not immediately clear how many individuals were erroneous. The erroneous families, CG848, CG755, HR8 and HR4 consisted of six, four, five and five individuals respectively. Within family CG755, errors could be accounted for by the fact that two of the individuals (Id. 7991 and Id. 8686) were more likely to originate

from the Southeastern Tasmanian and Furneaux races respectively, than from the breeding population classified race of the family (Furneaux) (Table 4.5). Likewise, family HB8 includes one individual (Id. 5507) that is likely to originate from the Strzelecki Ranges as opposed to its classified race of Southeastern Tasmania (Table 4.5). However, when this individual was dropped from the family, MLTR V. 3.0 (Ritland 2002) analysis indicated that the three remaining individuals still could not share one parent.

Principal coordinate analysis did not provide any further resolution of errors in the case of Family CG755 with no useful patterns of clustering occurring over the first three principal coordinate axes (Figure 4.6). Within family HB8, individual 5507, that was more likely to originate from a different race (STRUCTURE analysis), was clearly separated from the other three individuals along principal coordinate axis 1 (Figure 4.5). The other erroneous individual within this family could not be identified however. Individual 6497 was clearly separated from the remaining four members of family HR4 along principal coordinate axis one. MLTR V. 3.0 (Ritland 2002) analysis of the remaining individuals suggested that they could share a common parent without error. In the case of family CG848, no definitive resolution of the error was achieved. It was not possible in this case to test for common parentage of only two individuals.

Clonal errors between ramets sharing the same genotype code were found in only one out of the six cases tested. Four clones of Id. 5597 were genotyped, displaying a large number of allelic mismatches. Two grafts actually shared the same genotype with the remaining two grafts sharing another genotype in common.

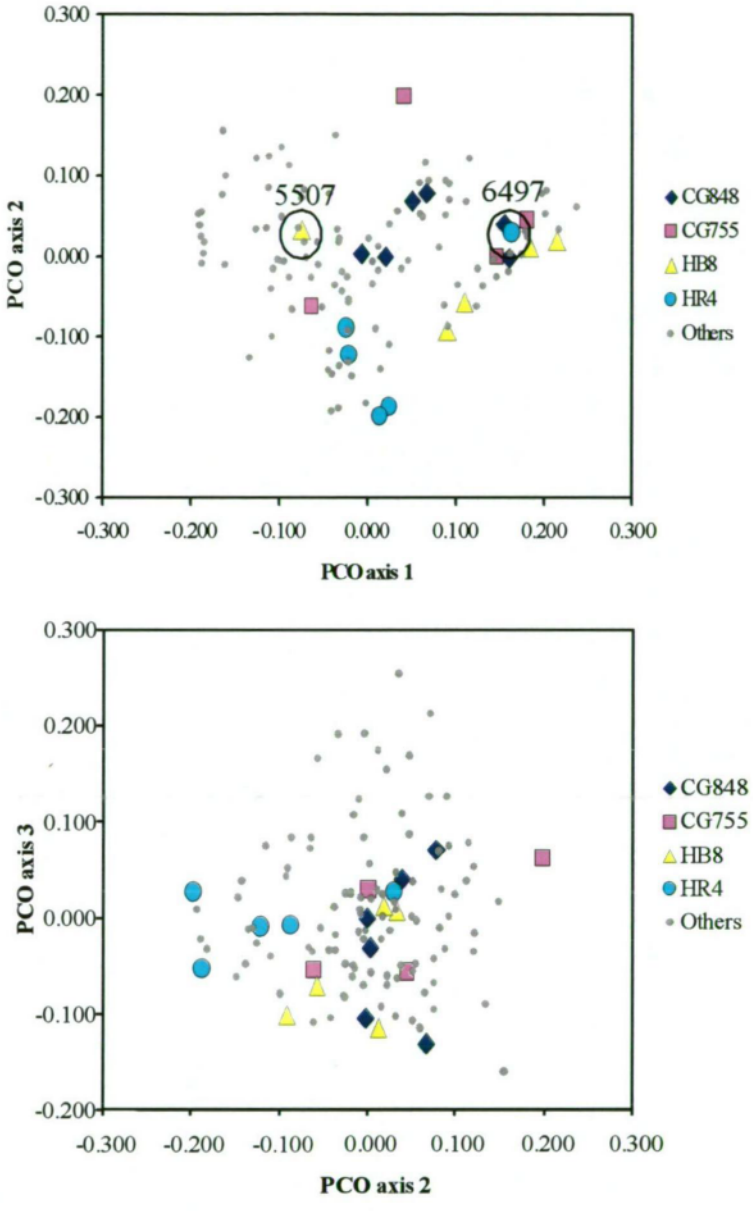



Figure 4.6: Principal coordinate analysis of the four erroneous families within the breeding population sample was carried out in an attempt to separate wrongly classified individuals from the remainder of the family. The first three principal coordinate axes (accounting for 7.0%, 3.8% and 3.6% of the total variation respectively) are plotted, with the four erroneous families indicated in the legend. The analysis also included all other individuals (indicated by grey dots). The circled individual 5507 was more likely to originate from a different race than that of the family classification. The circled individual 6497 was a likely erroneous individual classified into family HR4, but from the same race.

## Discussion

The overall levels of genetic diversity and heterozygosity within the Australian *E. globulus* breeding population sample ($H_e$ = 0.82, $H_o$ = 0.71) was within a similar range to that observed with microsatellite markers in other eucalypt breeding populations and seed orchard samples (Table 4.1). The levels of diversity within the native population sample, covering 10 *E. globulus* races, used in this study (340 individuals, $H_e$ = 0.81, $H_o$ = 0.66) were very similar to those already found in a smaller analysis (78 individuals, $H_e$ = 0.80, $H_o$ = 0.64) previously undertaken by Jones *et al.* (2002), based on an almost identical range of microsatellite markers (Steane *et al.* in prep). As is often the case in forest species (Kremer 1994), the native distribution of *E. globulus* is relatively genetically diverse (Steane *et al.* in prep). The expected and observed heterozygosity of the native distribution of the species, based on microsatellite markers, is comparable to that found in other native populations of eucalypts (Table 4.6), and a range of other forest trees including *Melaleuca alternifolia* Cheel (Rossetto *et al.* 1999), *Quercus petraea* Liebl, and *Quercus robur* L. (Streiff *et al.* 1998), *Camellia japonica* L. (Ueno *et al.* 2000), *Caryocar brasilense* Camb. (Collevatti *et al.* 1999), *Fagus crenulata* (Asuka *et al.* 2004) and *Symphonia globulifera* (Aldrich *et al.* 1998).

Table 4.6: Genetic diversity in breeding and natural populations of various eucalypts, based on microsatellite markers. Expected ($H_e$) and observed ($H_o$) heterozygosity are shown for breeding population samples (including seed orchard studies) (B) and native population samples (N).

| Species | Breeding or native population sample | $H_e$ | $H_o$ | Author |
|---|---|---|---|---|
| *E. globulus* | B | 0.82 | 0.71 | Present study |
| *E. dunnii* | B | 0.78 | 0.69 | Marcucci Poltri *et al.* 2003 |
| *E. grandis* | B | 0.90 | 0.78 | Jones *et al.* unpublished |
| *E. grandis* | B | 0.83 | 0.56 | Brondani *et al.* 1998 |
| *E. grandis* | B | 0.76 | 0.66 | Chaix *et al.* 2003 |
| *E. urophylla* | B | 0.86 | 0.59 | Brondani *et al.* 1998 |
| *E. globulus* | N | 0.81 | 0.66 | Present study |
| *E. globulus* | N | 0.80 | 0.64 | Jones *et al.* 2002 |
| *E. nitens* | N | 0.83 | 0.58 | Byrne *et al.* 1996 |
| *E. vernicosa* | N | 0.86 | 0.66 | McGowen et al. 2001 |
| *E. consideniana* | N | 0.60-0.62 | 0.59-0.61 | Glaubitz *et al.* 2003 |
| *E. sieberi* | N | 0.87 | 0.84 | Glaubitz et al. 2001 |
| *E. curtisii* | N | 0.54 | 0.47 | Smith *et al.* 2002 |
| *E. populnea/E. brownii* complex | N | 0.72-0.94 | 0.47-0.82 | Holman *et al.* 2003 |

The overall genetic diversity, based on expected and observed heterozygosity ($H_e$ = 0.82, $H_o$ = 0.71), within the Australian base breeding population sample was higher than that observed in both the native distribution sample that only included corresponding races of the breeding population sample ($H_e$ = 0.80, $H_o$ = 0.66), and the entire sample of the native distribution of the species (10 races, $H_e$ = 0.80, $H_o$ = 0.66). While not always the case, in species such as *Acacia aulacocarpa* (McGranahan *et al.* 1997), *E. pellita* (House and Bell 1996), *Picea glauca* (Rajora 1999), *Picea abies* (Bergmann and Ruetz 1991), *Pinus radiata,* (Moran and Bell 1987), it is common to see very little reduction in overall genetic diversity in forest tree breeding populations, seed orchards or the regions from where domestication of tree species has been based, compared to the native distribution of the species (Johnson and Lipow 2002; Lefèvre 2004). This has been the case in a wide range of species, including *Acacia mangium* (Butcher *et al.* 1998a; Butcher *et al.* 1998b), *Acacia auriculiformis* (Wickneswari and Norwati 1993), *Pseudotsuga menziesii* (El Kassaby and Ritland 1996a; El Kassaby and Ritland 1996b), *Pinus banksiana* (Godt *et al.* 2001), *Pinus caribea* var *caribaea* (Zheng and Ennos 1999), *Pinus palustris* (Schmidtling and Hipkins 1998), *Pinus radiata* (Moran and Bell 1987), *Pinus sylvestris* (Adams 1981; Szmidt and Muona 1985), one study of *Picea glauca* (Godt *et al.* 2001), *Picea glauca* x *engelmanni* (Stoehr and Elkassaby 1997), *Picea mariana* (Knowles 1985) and *Picea sitchensis* (Yeh and El-Kassaby 1980; Chaisurisri and El-Kassaby 1994).

While abundant overall genetic diversity may occur, measured with $H_o$ or $H_e$ (observed, expected heterozygosity) and $F_{st}$ (Nei 1987), breeding populations may show a significant reduction in allelic diversity (specifically rare alleles) relative to the native distribution of the species, due to the bottleneck effect of reduced effective population size associated with establishing the breeding population (Leberg 1992; Barker 2001). In other words, genetic bottlenecks due to small population size have a much greater effect on allelic richness than expected heterozygosity (Nei *et al.* 1975; Leberg 1992; Barker 2001). This appears to be the case in species such as *Picea glauca, Pinus banksiana* (Godt *et al.* 2001) and *Pseudotsuga menziessii* (El Kassaby and Ritland 1996a; El Kassaby and Ritland 1996b) among others, and is evident within the present study. Both native population samples of *E. globulus* had more private alleles in

comparison to the Australian breeding population even when variation in sample size was taken into account, with these private alleles being present in low frequencies (rare). Overall allelic richness, standardized to account for variation in sample size, was lower in the breeding population ($R_t$ = 14.26) than the two samples of the native distribution ($R_t$ = 16.31, 16.72). Alleles associated with functional genes that are of no current identified use to the breeding program may play a role in the future (Yanchuk 2001). In fact, there have been some situations in current tree breeding programs where rare alleles have been turned out to be highly desirable within the program. For example, the MGR gene for blister rust resistance in *Pinus lambertiana* has an overall allele frequency of 0.022 within the species (Kinloch 1992), necessitating the inclusion of a large number of individuals that express the gene to be included in the program to maintain that allele alone (Yanchuk 2001). Another highly desirable rare allele affects lignin properties in *Pinus taeda* (Ralph *et al.* 1997). Optimal maintenance of adaptive potential of a species and optimizing the genetic resource from a breeding point of view should include strategies to ensure the conservation of rare alleles (Namkoong *et al.* 1988; Yanchuk 2001).

Most interestingly, the observed heterozygosity was higher in the breeding population sample than the native distribution samples (0.71 versus 0.66) and the level of inbreeding within the breeding population, as measured by $F_{is}$, was around half that observed in the native distribution samples (0.059 versus 0.118 and 0.128). This may be an effect of spatial structure, however, the impact of inbreeding depression on the species (Hardner and Potts 1995, Tilyard *et al.* unpublished) may result in outcrossed families and individuals displaying better growth in progeny trials (e.g. Burgess 1996; Hardner *et al.* 1996) and hence, being selected into the breeding population. A range of species display an increase in fitness associated with heterozygosity (e.g. Charlesworth and Charlesworth 1987 and Hansson and Westerberg 2002). This indicates that within this program, the outcrossing potential of a parent may be biasing the prediction of the genetic worth of that family, with higher than expected outcrossing effectively inflating the estimated genetic value of a family.

Genetic diversity can be maintained in advanced generations by prioritizing genetic diversity in addition to genetic gain during selections, or by infusing material from other

breeding programs or from native populations into the breeding program (Namkoong *et al.* 1988; Hamann *et al.* 2004). From a tree breeding point of view, *in situ* native population conservation may provide an important resource of genetic diversity for infusion into breeding populations (Lefèvre 2004). An example of how long term maintenance of diversity within the wild type of a domesticated species benefits domestication can be seen in wheat, where the ancestral wild type is now viewed as useful source for broadening the genetic base of the "elite wheat breeding germplasm" (Reif *et al.* 2005). It is at this point that tree breeding and conservation of native forests share similar aims (Lefèvre 2004).

In the case of *E. globulus*, the Southern Tree Breeding Association have adopted a "rolling front strategy" with simultaneous evaluation of the base breeding population and advanced generation selections (Borralho and Dutkowski 1996; Borralho and Dutkowski 1998), which allows incorporation of more material from the base generation if the breeding values of this generation increase with more information or changing market requirements. This approach is likely to prevent further loss of rare alleles and maintain the genetic diversity in the breeding population into the future. Should a reduction in diversity occur in advanced generations of the breeding program, infusion of material from a wide range of sources is possible including base population trials, other *E. globulus* breeding programs throughout the world, the large pool of genetic diversity in native stands and overseas land races. Large patches of the native range of *E. globulus* are intact and many populations are in coastal and National Park Reserves, however, a native stand gene pool conservation strategy is yet to be developed (Potts *et al.* 2004).

Very little differences were seen in genetic diversity within the various races of *E. globulus* included in the breeding population sample. The King Island race displayed slightly lower genetic diversity than the other races, indicative of a potential bottleneck due to the geographic isolation of the island (Steane *et al.* in prep). As is often the case in forest populations (Kremer 1994), genetic differentiation was high between almost all of the geographic races within the breeding program (except South and Southeastern Tasmania and the Western and Eastern Otways, which are in close geographic proximity to each other). The hierarchical relationships between the seven races, as indicated by

UPGMA clustering, closely matched those seen in the more extensive native distribution sample containing the same races (Steane *et al.* in prep), with one exception of the King Island sample grouping more closely to the mainland races in the current study than to the eastern Tasmanian and Furneaux races seen in the study of Steane *et al.* (in prep). In addition, only two races within the breeding population sample were significantly genetically differentiated from the corresponding races of the native distribution sample (Southeastern Tasmania and Furneaux), however, the pair wise $F_{st}$ for these comparisons were still very low (0.054 and 0.021 respectively).

A large proportion of the genetic variation (80.6%) within the breeding populations is seen within the same open pollinated families. The AMOVA results suggest that like many native forest populations (Kremer 1994), significant variability resides within localities or populations (e.g. Maguire *et al.* 2002, dependent on the type of marker used). On the other hand, variation in racial origin of individuals within the breeding population accounted for only 4.9% of the total genetic variation within the breeding populations. This indicates that racial differences within *E. globulus* (Dutkowski and Potts 1999) only account for a small proportion of genetic variation within the species, indicated by neutral markers, while a large amount of genetic variation (89%) resides within populations, the majority of which, is actually within open pollinated families. This is consistent with high outcrossing and a large number of males contributing to open pollinated families as illustrated in Chapter 3. Bayesian analysis of population structure throughout the native distribution of the species (Steane *et al* in prep) supports the notion of less genetic variation between the classified races relative to the genetic diversity throughout the species.

Another issue in improving the accuracy of predictive breeding models is the accuracy of the pedigrees within the program. If individuals within seed orchards are misclassified to an erroneous race, locality, family or genotype, a direct impact on genetic assessment and breeding predictions will follow. Recently developed Bayesian assignment approaches applied to individual ancestry and population assignment (Cornuet *et al.* 1999; Rosenberg *et al.* 2003) and forensics (Foreman *et al.* 1997; Manel *et al.* 2002) provide an excellent opportunity to independently check for such errors in breeding populations. In the present study, nine individuals within the breeding

population shared a much higher probability of actually originating from a race other than that of their breeding program classification. In addition, the use of parentage analysis allowed the detection of erroneous individuals in breeding program classified families (Lexer *et al.* 2001; Ritland 2002; Jones and Ardren 2003), by checking that all individuals within the family share at least one common parent. In the current sample, four families were identified with errors within them (partially explained by the errors in racial classification in two families, clearly outlined in the results). Four individuals that were classified as clonal replicates of a single genotype actually consisted of a pair of two replicates of two distinct genotypes. Error in ramet classification within eucalypt breeding programs has been revealed in *E. nitens* (Keil and Griffin 1994; Vaillancourt *et al.* 1998), *E. grandis* (Keil and Griffin 1994) and *E. camaldulensis* (Keil and Griffin 1994) with the use of RAPD markers, and indicates that errors inherently occur at some stage during labeling, grafting, handling of seed and seedlings, planting of progeny trials and collection of material for testing, with potential economic consequences (Keil and Griffin 1994). While the studies just mentioned have succeeded in testing for clonal and within family pedigree errors, the current study has also provided a means of testing for less specific pedigree errors such as racial classification. Given that potential error at some level of classification was identified in around eight percent of the breeding population sample, consideration should be given firstly to the impact of this level of error on the predictions made by genetic models within the program, and secondly, to implementation of quality control procedures to reduce the probability of future error.

# Chapter 5: Concluding Discussion

### Major conclusions

The major conclusions that can be drawn from the evidence presented in this thesis are as follows:

1. Fine scale spatial genetic structure can exist within continuous populations of *Eucalyptus globulus*. In the Tinderbox Hills forest studied, a virtually linear decrease in genetic similarity with distance appears to occur between 10 and 100 m. In the mature cohort alone, a steeper reduction in the association between distance and genetic similarity was seen between trees of 0 to 55 m apart, providing support to the family patch structure hypothesised to occur in eucalypts (Eldridge *et al.* 1993) and shown to be the case in a remnant population of *E. globulus* (Skabo *et al.* 1998). Mean pairwise relatedness values between mature cohort trees that were separated by more than 45 m were above zero, providing further evidence for family group structure. Evidence of bi-parental inbreeding and/or full sibship was seen between trees separated by less than five meters (nearest neighbours). Levels of proximity based relatedness in the study site were significantly lower than that suggested to occur in the remnant *E. globulus* population studied by Skabo (1998).

2. Spatial genetic structure was due to a number of independent components of genetic variation spatially superimposed upon each other. A clinal pattern was revealed in the most informative component of variation within the population, with several other spatially structured independent genetic patches superimposed upon this. This multilayered pattern of genetic variation represents a very subtle, yet significant effect, with each independent component accounting for a small proportion of the total genetic variance within the population.

3. A weaker spatial genetic structure was displayed by the juvenile cohort than by the mature cohort. Assuming constant gene flow over time, this is most likely indicative of variation in founding demographics of the two cohorts. It is hypothesized that the mature cohort was established under more open forest

conditions, or after severe fire. However, a response to micro-site selection can not be excluded. The intermediate nature of the combined analysis suggests that parent – offspring relationships across cohorts are not occurring in close proximity.

4. A spatial shift between cohorts in the clinal pattern of genetic variation was revealed. This up slope shift followed the direction of the prevailing westerly and southwesterly winds at the site, and is a likely result of directional seed dispersal..

5. Effective gene flow was limited in the study population. The total effective gene flow, indicated by the position of established seedlings relative to their assigned parents in the mature cohort, displayed a leptokurtic decrease in frequency with increasing distance between parent and offspring. In this case, effective gene flow events over less than 40 m are over represented, and those above 40 m are under represented. This suggests a possible driving force behind the observed family structure within the population (Linhart 1989; Eldridge *et al.* 1993). While previously argued that family group structuring is primarily a product of limited seed dispersal (Eldridge *et al.* 1993; Skabo *et al.* 1998), the present study indicates that limited pollen dispersal is also a major contributing factor. Pollen dispersal displayed a leptokurtic reduction in frequency with increasing distance, with 37% of all genotyped seed pollinated by trees that were less than 15 m from the mother. Full siblings were more likely to be the products of mating between trees in close proximity to each other. On a wider scale, 61% of seed was assigned to parents separated by 50 m or less.

6. Bi-parental inbreeding is clearly occurring in open pollinated progeny, and there is some evidence for the survival of the products of these matings to reproductive maturity. Bi-parental inbreeding appears to result from the combination of significant spatial genetic structure and limited gene flow. The level of self pollination was comparable to the level of bi-parental inbreeding in the population, however, self pollination is clearly the dominant cause of inbreeding depression in open pollinated progenies. There is also evidence for the rare

establishment of selfed progeny into the juvenile cohort of the native forest. The survival of these progeny is surprising when considering the level of inbreeding depression predicted to occur in inbred progeny (Tilyard *et al.* unpublished; Hardner and Potts 1995; Hardner *et al* 1998). Nevertheless, survival of inbred progeny and restricted gene flow does not appear to be sufficient to perpetuate spatial genetic structure nor reduce heterozygosity over time. It is therefore likely that directional, as opposed to random gene flow, may play a major role in maintaining genetic diversity within the population.

7. The Australian national *E. globulus* breeding population sample displayed a high level of genetic heterozygosity. Expected and observed heterozygosity levels were actually higher than that within a wide ranging sample of the native distribution of the species made by Steane *et al.* (in prep). An increase in observed heterozygosity suggests that artificial selection has favoured families with high outcrossing rates. However, allelic richness was lower in the breeding population, indicating the need for careful future management of the program to maintain genetic diversity.

8. Despite statistically significant genetic variation between races, localities and families, the major source of genetic variation within the breeding population occurs between individuals of the same open pollinated family. This is consistent with the high levels of outcrossing and heterozygosity in the native forest and the large number of fathers represented in the open pollinated progenies.

9. A significant proportion of pedigree error was detected in the Australian national *E. globulus* breeding population. Specific errors in the race classification of at least seven individuals out of the 149 studied were identified, with two classified families containing at least one individual that did not share a common parent with the remainder of the family. Four classified ramets of the same genotype actually consisted of two ramets of two different genotypes. This clearly illustrates the applicability of molecular markers to pedigree testing and quality control in tree breeding programs (Vaillancourt *et al.* 1998).

**Future directions**

The application of current genetic analysis techniques have clearly illustrated that the spatial and temporal genetic variation predicted to occur in native forest by classic population genetic theory (e.g. Hartl and Clarke, 1989; Malécot, 1948; Wright, 1951) do occur in this continuous native population of *E. globulus*. Infact, this study has revealed a complex and dynamic picture of genetic interactions at an incredibly fine scale. At the local population level, a number of subtle genetic dynamics that complicate the underlying concept of isolation by distance (Wright 1943) are evident. However, with every insight into these processes comes an abundant suite of subsequent questions. Interpretation of the spatial and temporal genetic variation described in this study has focused on neutral gene flow dynamics. The influence of micro–scale selection within the site remains unspecified, and can only be revealed with a detailed investigation of possible environmental correlates. The application of maternal markers such as chloroplast microsatellites (Provan *et al.* 2001; Steane *et al.* 2005) will provide a much more accurate quantification of seed dispersal than that predicted by the comparison of pollen dispersal and total effective gene flow. While this study has advanced the understanding of spatial genetic dynamics on a local scale, the next step is to understand the longer distance component of gene flow (e.g. Smouse and Sork 2004; Smouse *et al.* 2001). In combination with the parameters presented and those detailed by Hardner and Potts (1995), Hardner *et al* (1996; 1998), Patterson *et al.* (2004), Steane *et al.* (in prep) and McKinnon *et al.* (2004), the predictive modeling of population genetic and evolutionary processes (e.g. Austerlitz *et al.* 2003) of the species will be possible.

# References


Adams WT (1981) Population genetics and gene conservation in Pacific Northwest conifers. In 'Evolution today. Proceedings of the Second International Congress on Systematic and Evolutionary Biology'. (Eds GGE Scudder and JL Reveal) pp. 401-405

Adams WT (1992) Gene dispersal within forest tree populations. *New Forests* **6**, 217-240.

Adams WT, Birkes DS (1991) Estimating mating patterns in forest tree populations. In 'Biochemical markers in the population genetics of forest trees'. (Eds S Fineschi, ME Malvolti, F Cannata and HH Hattemer) pp. 157-172. (SPB Academic Publishing: The Hague)

Aldrich PR, Hamrick JL, Chavarriaga P, Kochert G (1998) Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. *Molecular Ecology* **7**, 933-944.

Araújo JA, Sousa R, Lemos L, Borralho NMG (1996) Estimates of genetic parameters and prediction of breeding values for growth in *Eucalyptus globulus* combining clonal and full-sib progeny trial information. *Silvae Genetica* **45**, 223-226.

Ardlie KG, Lunetta KL, Seielstad M (2002) Testing for population subdivision and association in four case-control studies. *American Journal of Human Genetics* **71**, 304-311.

Arnold S (1994) Bateman's principles and the measurement of sexual selection in plants and animals. *American Naturalist* **144**, 126-149.

Ashton DH (1981) Fire in tall open forests (wet sclerophyll forests). In 'Fire and the Australian biota'. (Ed. WH Groves))

Ashton DH, Sandiford LM (1988) Natural hybridisation between *Eucalyptus regnans* F. Muell. and *E. macrorhyncha* F. Muell. in the Cathedral Range, Victoria. *Australian Journal of Botany* **36**, 1-22.

Astorga R, Soria F, Basurco F, Toval G (2004) Diversity analysis and genetic structure of *Eucalyptus globulus* Labill. In 'Eucalyptus in a changing world'. Aveiro, Portugal. (Eds NMG Borralho, JS Pereira, C Marques, J Coutinho, M Madeira and M Tomé) pp. 351-358. (IUFRO)

Asuka Y, Tomaru N, Nisimura N, Tsumura Y, Yamamoto S (2004) Heterogeneous genetic structure in a *Fagus crenata* population in an old-growth beech forest revealed by microsatellite markers. *Molecular Ecology* **13**, 1241-1250.

Austerlitz F, Dick CW, Dutech C, Klein EK, Oddou-Muratorio S, Smouse PE, Sork VL (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* **13**, 937-954.

Austerlitz F, Garnier-Géré PH (2003) Modelling the impact of colonisation on genetic diversity and differentiation of forest trees: interaction of life cycle, pollen flow and seed long-distance dispersal. *Heredity* **90**, 282-290.

Austerlitz F, Smouse PE (2001) Two-generation analysis of pollen flow across a landscape. II. Relation between phift, pollen dispersal and inter-female distance. *Genetics* **157**, 851-857.

Austerlitz F, Smouse PE (2002) Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter. *Genetics* **161**, 355-363.

Avise JC (1994) 'Molecular markers, natural history and evolution.' (Chapman and Hall: New York)

Balanovskaya EV, Nurbaev SD (1997) A computer-aided method for gene-geographic study of a population gene pool: populations in the space of principal components. *Genetika* **33**, 1703-1718.

Balding DJ (2003) Likelihood-based inference for genetic correlation coefficients. *Theoretical Population Biology* **63**, 221-230.

Barber HN (1965) Selection in natural populations. *Heredity* **20**, 551-572.

Barber HN, Jackson WD (1957) Natural selection in action in *Eucalyptus*. *Nature* **179**, 1267-1269.

Barbour RC, Potts BM, Vaillancourt RE (2005) Pollen dispersal from exotic eucalypt plantations. *Conservation Genetics* **6**, 253-257.

Barker JSF (2001) Conservation and management of genetic diversity: a domestic animal perspective. *Canadian Journal of Forest Research* **31**, 588-595.

Bateman AJ (1948) Intrasexual selection in *Drosophila*. *Heredity* **2**, 249-368.

Beaumont M (1999) Detecting population expansion and decline using microsatellites. *Genetics* **153**, 2013-2029.

Bergmann F, Ruetz W (1991) Isozyme genetic variation and heterozygosity in random tree samples and selected orchard clones from the same Norway spruce populations. *Forest Ecology and Management* **46**, 39-47.

Berry O, Tocher MD, Sarre SD (2004) Can assignment tests measure dispersal? *Molecular Ecology* **13**, 551-561.

Blouin MS (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution* **18**, 503-511.

Blouin MS, Parsons M, Lacaille V, Lotz S (1996) Use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology* **5**, 393-401.

Boland DJ, Brooker MIH, Turnbull JW, Kleinig DA (1980) '*Eucalyptus* seed.' (CSIRO: Melbourne)

Borralho NMG, Dutkowski GW (1996) A 'rolling-front' strategy for breeding trees. In 'Tree Improvement for Sustainable Tropical Forestry'. Caloundra, Queensland. (Eds MJ Dieters, AC Matheson, DG Nikles, CE Harwood and SM Walker) pp. 317-322. (Queensland Forest Research Institute)

Borralho NMG, Dutkowski GW (1998) Comparison of rolling front and discrete generation breeding strategies for trees. *Canadian Journal of Forest Research* **28**, 987-993.

Borralho NMG, Potts BM (1996) Accounting for native stand characteristics in genetic evaluations of open pollinated progeny from *Eucalyptus globulus* base population. *New Forests* **11**, 53-64.

Bradshaw AD (1972) Some of the evolutionary consequences of being a plant. *Evolutionary Biology* **5**, 25-47.

Bretting PK, Duvick DN (1997) Dynamic conservation of plant genetic resources. *Advances in Agronomy* **61**, 1-51.

Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theoretical and Applied Genetics* **97**, 816-827.

Brooker MIH (2000) A new classification of the genus *Eucalyptus* L'Her. (Myrtaceae). *Australian Systematic Botany* **13**, 79-148.

Brown AHD, Matheson AC, Eldridge KG (1975) Estimation of the mating system of *Eucalyptus obliqua* L'Herit. by using allozyme polymorphisms. *Australian Journal of Botany* **23**, 931-949.

Bundock PC, Hayden M, Vaillancourt RE (2000) Linkage maps of *Eucalyptus globulus* using RAPD and microsatellite markers. *Silvae Genetica* **49**, 223-232.

Burczyk J, Adams WT, Moran GF, Griffin AR (2002) Complex patterns of mating revealed in a *Eucalyptus regnans* seed orchard using allozyme markers and the neighbourhood model. *Molecular Ecology* **11**, 2379-2391.

Burczyk J, Adams WT, Shimizu JY (1996) Mating patterns and pollen dispersal in a natural knobcone (*Pinus attenuata* Lemmon) stand. *Heredity* **77**, 251-260.

Burczyk J, Prat D (1997) Male reproductive success in *Pseudotsuga menziesii* (Mirb.) Franco: the effects of sptatial structure and flowering characteristics. *Heredity* **79**, 638-647.

Burley J (2001) Genetics in sustainable forestry: the challenges for forest genetics and tree breeding in the new millennium. *Canadian Journal of Forest Research* **31**, 561-565.

Butcher PA, Moran GF, De Croocq S (1998a) Use of molecular markers in domestication and breeding programs for acacias. In 'ACIAR Proceedings No. 82'. (Eds JW Turnbull, HR Crompton and Pinyopusarerk) pp. 211-215. (ACIAR: Canberra)

Butcher PA, Moran GF, Perkins HD (1998b) RFLP diversity in the nuclear genome of *Acacia mangium*. *Heredity* **81**, 205-213.

Byrne M, Parrish TL, Moran GF (1998) Nuclear RFLP diversity in *Eucalyptus nitens*. *Heredity* **81**, 225-233.

Caballero A, Santiago E (1994) Prediction of effective size of populations under selection and non-random mating. In 'Proceedings of the 5th World Congress on Genetics Applied to Livestock Production'. University of Guelph pp. 143-146. (University of Guelph)

Campinhos EN, Peters-Robinson I, Bertolucci FL, Alfenas AC (1998) Interspecific hybridisation and inbreeding effect in seed from a *Eucalyptus grandis* x *E. urophylla* clonal orchard in Brazil. *Genetics and Molecular Biology*.

Caughley G (1994) Directions in conservation biology [Review]. *Journal of Animal Ecology* **63**, 215-244.

Cavalli-Sforza LL (1966) Population structure and human evolution. *Proceedings of the Royal Society of London, Series B* **164**, 362-379.

Cavalli-Sforza LL (2001) 'Genes, people and languages.' (Penguin Books: London)

Cegelski CC, Waits LP, Anderson NJ (2003) Assessing population structure and gene flow in Montana wolverines (*Gulo gulo*) using assignment-based approaches. *Molecular Ecology* **12**, 2907-2918.

Chaisurisri K, El-Kassaby YA (1994) Genetic diversity in a seed production population vs natural populations of Sitka spruce. *Biodiversity and Conservation* **3**, 512-523.

Chaix G, Gerber S, Razafimaharo V, Vigneron P, Verhaegan D, Hamon S (2003) Gene flow estimation with microsatellites in a Malagasy seed orchard of *Eucalyptus grandis*. *Theoretical & Applied Genetics* **107**, 705-712.

Chakraborty R, Meagher TR, Smouse PE (1988) Parentage analysis with genetic markers in natural popuations. 1. The expected proportion of offspring with unambiguous paternity. *Genetics* **118**, 527-536.

Chakraborty R, Shaw M, Schull WJ (1974) Exclusion of paternity: the current state of the art. *American Journal of Human Genetics* **26**, 477-488.

Charlesworth D, Charlesworth B (1997) Inbreeding depression and its evolutionary consequences. *Annual Review of Ecological Systematics* **18**, 237-268.

Chung MG, Chung MY, Oh GS, Epperson BK (2000) Spatial genetic structure in a *Neolitsea sericea* population (Lauraceae). *Heredity* **85**, 490-497.

Collevatti RG, Brondani RV, Grattapaglia D (1999) Development and characterisation of microsatellite markers for genetic analysis of a Brazilian endangered species *Caryocar brasiliense*. *Heredity* **83**, 784-756.

Cornuet J, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**, 1989-2000.

Cotterman CW (1954) Estimation of gene frequencies in nonexperimental populations. In 'Statistics and mathematics in biology'. (Eds O Kempthorne, TA Bancroft, JW Gowen and JL Lush) pp. 449-465. (Iowa State College Press: Ames)

Crawford TJ (1984) The estimation of neighbourhood parameters for plant populations. *Heredity* **52**, 273-283.

Cremer KW (1966) Dissemination of seed from *Eucalyptus regnans*. *Australian Forestry* **30**, 33-37.

Cremer KW (1977) Distance of seed dispersal in eucalypts estimated from seed weights. *Australian Forest Research* **7**, 225-228.

Crow JF (1986) 'Basic concepts in population, quantitative and evolutionary genetics.' (W. H. Freeman and Company: New York)

Davies N, Villablanca FX, Roderick GK (1999) Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends in Ecology & Evolution* **14**, 17-21.

Degen B, Caron H, Bandou E, Maggia L, Chevallier MH, Leveau A, Kremer A (2001) Fine-scale spatial genetic structure of eight tropical tree species as analysed by RAPDs. *Heredity* **87**, 497-507.

Devlin B, Ellstrand NC (1990) The development and application of a refined method for estimating gene flow from angiosperm paternity analysis. *Evolution* **44**, 248-259.

Devlin B, Roeder K, Ellstrand NC (1988) Fractional paternity assignment: theoretical development and comparison to other methods. *Theoretical and Applied Genetics* **76**, 369-380.

Di Giacomo F, Luca F, *et al.* (2003) Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. *Molecular Phylogenetics and Evolution* **28**, 387-395.

Dick CW, Etchelecu G, Austerlitz F (2003) Pollen dispersal of tropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest. *Molecular Ecology* **12**, 753-764.

Doebley J (1989) Isozymic evidence and evolution of crop plants. In 'Isozymes in plant biology'. (Eds ED Soltis and PM Soltis) pp. 165-191. (Oregon Dioscordes: Portland)

Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. *Nature* **386**, 485-488.

Doganlar S, Frary A, Daunay M-C, Lester RN, Tanksley SD (2002) Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics* **161**, 1713-1726.

Doligez A, Joly HI (1997) Genetic diversity and spatial structure within a natural stand of a tropical forest tree species, *Carapa procera* (Meliaceae), in French Guiana. *Heredity* **79**, 72-82.

Double MC, Cockburn A, Barry SC, Smouse PE (1997) Exclusion probabilities for single-locus paternity analysis when related males compete for matings. *Mol Ecol* **6**, 1155-1166.

Dow BD, Ashley MV (1998) High levels of gene flow in Bur Oak revealed by paternity analysis using microsatellites. *Journal of Heredity* **89**, 62-70.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* **12 (1)**, 13-15.

Duchesneau R, Morin H (1999) Early seedling demography in balsam fir seedling banks. *Canadian Journal of Forest Research* **29**, 1502-1509.

Duncan F (1981) Regeneration and species diversity in a Tasmanian dry sclerophyll forest. In 'Proceedings of the Symposium on Fire and Forest Management, May 1981'. Hobart. (Tasmanian Conservation Trust)

Duncan F, Brown MJ (1985) 'Dry sclerophyll vegetation in Tasmania. Extent and conservation status of the communities. Wildlife Div. Tech. Rep. 85/1. NPWS, Tasmania.'

Dutkowski GW, Potts BM (1999) Geographical patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. *Australian Journal of Botany* **46**, 237-263.

Dutkowski GW, Potts BM, Borralho NMG (1997) Revised racial classification of *Eucalyptus globulus* ssp. *globulus* and the importance of including race in analysis of progeny trials. In 'IUFRO Conference on Silviculture and Improvement of Eucalypts'. Salvador, Bahia, Brasil pp. 322-329. (EMBRAPA, Colombo)

Ehrlich PR, Raven PH (1969) Differentiation of populations. *Science* **165**, 1228-1232.

El Kassaby YA, Ritland K (1996a) Genetic variation in low elevation Douglas-fir of British Columbia and its relevance to gene conservation. *Biodiversity and Conservation* **5**, 779-794.

El Kassaby YA, Ritland K (1996b) Impact of selection and breeding on the genetic diversity in Douglas-fir. *Biodiversity and Conservation* **5**, 795-813.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L) Skeels] endemic to Morocco. *Theoretical and Applied Genetics* **92**, 832-839.

Eldridge K, Davidson J, Harwood C, van Wyk G (1993) 'Eucalypt Domestication and Breeding.' (Clarendon Press: Oxford)

Ellis MF, Sedgley M (1993) Gynodioecy and male sterility in *Eucalyptus leucoxylon* F. Muell. (Myrtaceae). *International Journal of Plant Sciences* **154**, 314-324.

Ellstrand NC (1992a) Gene flow among seed plant populations. *New Forests* **6**, 241-256.

Ellstrand NC (1992b) Gene flow by pollen: implications for plant conservation genetics. *Oikos* **63**, 77-86.

Endler JA (1973) Gene flow and population differentiation. *Science* **179**, 243-250.

Endler JA (1977) 'Geographical variation, speciation and clines.' (Princeton: Princeton University Press)

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72**, 250-259.

Epperson BK (1992) Spatial structure of genetic variation within populations of forest trees. *New Forests* **6**, 257-278.

Epperson BK (1995a) Fine-scale spatial structure - Correlations for individual genotypes differ from those for local gene-frequencies. *Evolution* **49**, 1022-1026.

Epperson BK (1995b) Spatial Distributions of Genotypes under Isolation by Distance. *Genetics* **140**, 1431-1440.

Epperson BK (2000) Spatial genetic structure and non-equilibrium demographics within plant populations. *Plant Species Biology* **15**, 269-279.

Epperson BK, Alvarez Buylla ER (1997) Limited seed dispersal and genetic structure in life stages of *Cecropia obtusifolia*. *Evolution* **51**, 275-282.

Epperson BK, Li TQ (1996) Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 10528-10532.

Eriksson G (2001) Conservation of the noble hardwoods in Europe. *Canadian Journal of Forest Research* **31**, 577-587.

Eriksson G, Namkoong G, Roberds JH (1993) Dynamic gene conservation for uncertain futures. *Forest Ecology and Management* **62**, 15-37.

Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11**, 1591-1604.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Excoffier L (2001) Analysis of population subdivision. In 'Statistical genetics'. (Eds DJ Balding, M Bishop and C Cannings) pp. 271-307. (John Wiley and Sons: Chichester)

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances around DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Fagundes NJR, Bonatto SL, Callegari-Jacques SM, Salzano FM (2002) Genetic, geographic, and linguistic variation among South American Indians: Possible sex influence. *American Journal of Physical Anthropology* **117**, 68-78.

Fenster CB (1991) Gene flow in *Chamaecrista fasciculata* (Leguminosae). 2. Gene establishment. *Evolution* **45**, 410-422.

Fenster CB, Vekemans X, Hardy OJ (2003) Quantifying gene flow from spatial genetic structure data in a metapopulation of *Chamaecrista fasciculata* (Leguminosae). *Evolution* **57**, 995-1007.

Fins L, Friedman ST, Brotschol JV (Eds) (1992) 'Handbook of quantitative forest genetics.' (Kluwer Academic Publishers: Dordrecht)

Fisher RA (1950) Gene frequencies in a cline determined by selection and diffusion. *Biometrics* **6**, 353-361.

Foreman LA, Smith AFM, Evett IW (1997) Bayesian analysis of DNA profiling data in forensic identification applications. *Journal of the Royal Statistical Society Series a-Statistics in Society* **160**, 429-459.

Futuyama D (1998) 'Evolutionary biology.' (Sinauer Associates: Sunderland)

Gaut BS, Le Thierry d'Ennequin M, Peek AS, Sawkins MC (2000) Maize as a model for the evolution of plant nuclear genomes. *PNAS* **97**, 7008-7015.

Gemas VJV, Neves LO, Araújo C, Fevereiro P (2004) How can ISSR markers be useful in the management of genetic resources in a *Eucalyptus globulus* ssp. *globulus* baseline collection. In 'Eucalyptus in a changing world'. Aveiro, Portugal. (Eds NMG Borralho, JS Pereira, C Marques, J Coutinho, M Madeira and M Tomé) pp. 373-377. (IUFRO)

Gerber S, Chabrier P, Kremer A (2003) FAMOZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes* **3**, 479-481.

Gerber S, Mariette S, Streiff R, Bodénès C, Kremer A (2000) Comparison of microsatellite and amplified fragment length polymorphism markers for parentage analysis. *Molecular Ecology* **9**, 1037-1048.

Gill AM (1997) Eucalypts and fires: interdependent or independent? In 'Eucalypt ecology: individuals to ecosystems'. (Eds J Williams and J Woinarski) pp. 151-167. (Cambridge University Press: Cambridge)

Gizachew B (2002) ' *Eucalyptus globulus* provenance/progeny trial and *Eucalyptus camaldulensis* progeny trial in Ethiopia.' Ethiopian Agricultural Research Organisation, Forestry Research Centre, Canberra.

Glaubitz JC, Moran GF (2000) Genetic tools: the use of biochemical and molecular markers. In 'Forest conservation genetics: principals and practice'. (Eds A Young, D Boshier and T Boyle) pp. 39-59. (CSIRO Publishing: Collingwood)

Gliddon C, Belhassen E, Gouyon PH (1987) Genetic neighbourhoods in plants with diverse systems of mating and different patterns of growth. *Heredity* **59**, 29-32.

Godt MJW, Hamrick JL, Edwards-Burke MA, Williams JH (2001) Comparisons of genetic diversity in white spruce (*Picea glauca*) and jack pine (*Pinus banksiana*) seed orchards with natural populations. *Canadian Journal of Forest Research* **31**, 943-949.

Goldstein DB, Linares AR, Cavallisforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463-471.

Goldstein DB, Schlötterer C (1999) 'Microsatellites: evolution and application.' (Oxford University Press: Oxford)

Gonzalez-Martinez SC, Gerber S, Cervera MT, Martinez-Zapater JM, Alia R, Gil L (2003) Selfing and sibship structure in a two-cohort stand of maritime pine (*Pinus pinaster* Ait.) using nuclear SSR markers. *Annals of Forest Sciences* **60**, 115-121.

Gonzalez-Martinez SC, Gerber S, Cervera MT, Martinez-Zapater JM, Gil L, Alia R (2002) Seed gene flow and fine-scale structure in a Mediterranean pine (*Pinus pinaster* Ait.) using nuclear microsatellite markers. *Theoretical and Applied Genetics* **104**, 1290-1297.

Gore PL, Potts BM (1995) The genetic control of flowering time in *Eucalyptus globulus*, *E. nitens* and their F1 hybrids. In 'Eucalypt Plantations: Improving Fibre Yield and Quality'. Hobart, Tasmania. (Eds BM Potts, NMG Borralho, JB Reid, RN Cromer, WN Tibbits and CA Raymond) pp. 241-242. (CRC for Temperate Hardwood Forestry)

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.

Goudet J, Raymond M, de-Meeus T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics* **144**, 1933-1940.

Grandillo S, Tanksley SD (1996) QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. *Theoretical and Applied Genetics* **92**, 935-951.

Griffin AR (1990) Effects of inbreeding on growth of forest trees and implication for management of seed supplies for plantation programmes. In 'Reproductive ecology of tropical forest plants'. (Eds KS Bawa and M Hadley) pp. 355-372. (UNESCO and The Parthenon Publishing Group: Paris, France)

Griffin AR, Cotterill PP (1988) Genetic variation in growth of outcrossed, selfed and open-pollinated progenies of *Eucalyptus regnans* and some implications for breeding strategy. *Silvae Genetica* **37**, 124-131.

Griffin AR, Moran GF, Fripp YJ (1987) Preferential outcrossing in *Eucalyptus regnans* F. Muell. *Aust J Bot* **35**, 465-475.

Grosser C, Vaillancourt RE, Potts BM, O'Sullivan H (2001) Paternity Analysis in a *Eucalyptus nitens* clonal seed orchard. In 'Developing the Eucalypt of the Future'. Valdivia, Chile. (INFOR)

Haldane JBS (1948) The theory of a cline. *Journal of Genetics* **48**, 277-284.

Hamann A, Aitken SN, Yanchuk AD (2004) Cataloguing *in situ* protection of genetic resources for major commercial forest trees in British Columbia. *Forest Ecology and Management* **197**, 295-306.

Hamrick J (1983) The distribution of genetic variation within and among natural plant populations. In 'Genetics and conservation'. (Eds CM Schonewald-Cox, SM Chambers, B MacBryde and L Thomas) pp. 335-348. (Benjamin-Cummings: London)

Hamrick JL (1982) Plant population genetics and evolution. *American Journal of Botany* **69**, 1685-1693.

Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In 'Plant population genetics, breeding and genetic resources'. (Eds AHD Brown, MT Clegg, AL Kahler and BS Weir) pp. 44-64. (Sinauer: Sunderland)

Hamrick JL, Loveless MD (1986) The influence of seed dispersal mechanisms on the genetic structure of plant populations. In 'Frugivores and seed dispersal'. (Eds A Estrada and TH Flemin). (Dr Junk Publications: Dordrecht, Netherlands)

Hamrick JL, Murawski DA, Nason JD (1993) The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. *Vegetatio* **108**, 281-297.

Hamrick JL, Nason JD (1996) Consequences of dispersal in plants. In 'Population dynamics in ecological space and time'. (Eds OE Rhodes, RK Chesser and MH Smith) pp. 203-236. (University of Chicago Press: Chicago)

Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JEO (2002) African pastoralism: genetic imprints of origins and migrations. *Science* **296**, 336-339.

Hanson WD (1966) Effects of partial isolation (distance), migration, and different fitness requirements among environmental pockets upon steady state gene frequencies. *Biometrics* **22**, 453-468.

Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology* **11**, 2467-2474.

Hardner CM (1996) Inbreeding in three forest eucalypts. Ph.D. thesis, University of Tasmania.

Hardner CM, Potts BM (1995) Inbreeding depression and changes in variation after selfing in *Eucalyptus globulus* ssp *globulus*. *Silvae Genetica* **44**, 46-54.

Hardner CM, Potts BM, Gore PL (1998) The relationship between cross success and spatial proximity of *Eucalyptus globulus* ssp. *globulus* parents. *Evolution* **52**, 614-618.

Hardner CM, Vaillancourt RE, Potts BM (1996) Stand density influences outcrossing rate and growth of open-pollinated families of *Eucalyptus globulus*. *Silvae Genetica* **45**, 226-228.

Hardy OJ, Gonzalez-Martinez SC, Colas B, Fréville H, Mignot A, Olivieri I (2004) Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae). II. Correlated paternity within and among sibships. *Genetics* **168**, 1601-1614.

Hardy OJ, Vanderhoeven S, Meerts P, Vekemans X (2000) Spatial autocorrelation of allozyme and quantitative markers within a natural population of *Centaurea jacea* (Asteraceae). *Journal of Evolutionary Biology* **13**, 656-667.

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**, 145-154.

Hardy OJ, Vekemans X (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618-620.

Harlan RJ (1992) 'Crops and man. Second edition.' (Crop Science Society of America, Inc.: Madison, Wisconsin)

Hartl DL, Clark AG (1989) 'Principles of population genetics, 2nd edition.' (Sinauer Associates: Sunderland)

Harwood C (1999) Domestication of Australian tree species for agroforestry. *Forest, Farm, and Community Tree Research* **Special issue**, 64-72.

Hedrick PW (1990) Mating Systems and Evolutionary Genetics.

Hedrick PW (2004) Recent developments in conservation genetics. *Forest Ecology and Management* **197**, 3-19.

Hernandez-Verdugo S, Luna-Reyes R, Oyama K (2001) Genetic structure and differentiation of wild and domesticated populations of *Capsicum annuum* (Solanaceae) from Mexico. *Plant Systematics and Evolution* **226**, 129-142.

Heywood JS (1991) Spatial analysis of genetic variation in plant populations. *Annual Review of Ecological Systematics* **22**, 335-355.

Hingston A, Potts BM, McQuillan PB (2004a) Pollination services provided by various size classes of flower visitors to *Eucalyputs globulus* ssp. *globulus* (Myrtaceae). *Australian Journal of Botany* **52**, 353-369.

Hingston A, Potts BM, McQuillan PB (2004b) The swift parrot, *Lathamus discolor* (Psittacidae), social bees (Apidae) and native insects as pollinators of *Eucalyptus globulus* ssp. *globulus* (Myrtaceae). *Australian Journal of Botany* **52**.

Hingston AB (2002) Pollination ecology of *Eucalyptus globulus* Labill. ssp. *globulus* and *E. nitens* (Deane & Maiden) Maiden (Myrtaceae). PhD thesis, University of Tasmania.

Hingston AB, Potts BM (1998a) Floral visitors of *Eucalyptus globulus* subsp. *globulus* in eastern Tasmania. *Tasforests* **10**, 125-140.

Hingston AB, Potts BM (1998b) 'A pictorial catalogue of insects collected from flowers of *Eucalyptus globulus* subsp. *globulus* in eastern Tasmania.' Cooperative Research centre for Sustainable Production Forestry, 5, Hobart, Tasmania, Australia.

Hodge GR, Volker PW, Potts BM, Owen JV (1996) A comparison of genetic information from open-pollinated and control-pollinated progeny tests in two eucalypt species. *Theoretical and Applied Genetics* **92**, 53-63.

Holman JE, Hughes JM, Fensham RJ (2003) A morphological cline in *Eucalyptus*: a genetic perspective. *Molecular Ecology* **12**, 3013-3025.

Holsinger KE (1999) Analysis of genetic diversity in geographically structured populations: a Bayesian perspective. *Hereditas* **130**, 245-255.

Holsinger KE, Lewis PO, Dey DK (2002) A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology* **11**, 1157-1164.

Holsinger KE, Wallace LE (2004) Bayesian approaches for the analysis of population genetic structure: an example from *Platanthera leucophaea* (Orchidaceae). *Molecular Ecology* **13**, 887-894.

Hopper SD, Moran GF (1981) Bird pollination and the mating system of *Eucalyptus stoatei*. *Australian Journal of Botany* **29**, 625-638.

House APN, Bell JC (1994) Isozyme variation and mating system in *Eucalyptus urophylla* S. T. Blake. *Silvae genetica* **43**, 167-176.

House APN, Bell JC (1996) Genetic diversity, mating system and systematic relationships in two red mahoganies, *Eucalyptus pellita* and *E. scias*. *Australian Journal of Botany*.

Infante LP, Prado DJA (1989) Growth of 35 provenances of *Eucalyptus globulus* spp. *globulus* in the nursery stage. *Ciencia e Investigacion Forestal* **7**.

Infante P, Prado JA (1991) Juvenile growth of 32 provenances and 203 families of *Eucalyptus globulus* subsp. *globulus* in the coastal zone of Region VIII of Chile. *Ciencia e Investigacion Forestal* **5**, 251-265.

Ipinza RH, García X, Apiolaza LA, Molina MP, Chung P, Parra P (1994) Variación juvenil de un ensayo de procedencias y familias de *Eucalyptus globulus* subsp. *globulus* Labill, en la séptima región, Chile. *Ecología* **8**, 259-270.

Irwin AJ, Hamrick JL, Godt MJW, Smouse PE (2003) A multiyear estimate of the effective pollen donor pool for *Albizia julibrissin*. *Heredity* **90**, 187-194.

Jackson WD (Ed.) (1981) 'The Vegetation of Tasmania.'

Jarne P, Lagoda P (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* **11**, 424-429.

Jarvis S, Borralho NMG (1995) 'The STBA cooperative breeding strategy of *Eucalyptus globulus* and *Eucalyptus nitens*.' Southern Tree Breeding Association, and Cooperative Research Centre for Sustainable Production Forestry, TR 95-03, Mount Gambier, South Australia.

Johnson R, Lipow S (2002) Compatibility of breeding for increased wood production and long-term sustainability: the genetic variation of seed orchard seed and associated risks. In 'Proceedings from the Wood Compatibility Initiative Workshop, number 18' pp. 169-179

Jones GJ, Ardren WR (2003) Methods of parentage analysis in natural populations. *Molecular Ecology* **12**, 2511-2523.

Jones RC, Steane DA, Potts BM, Vaillancourt RE (2002) Microsatellite and morphological analysis of *Eucalyptus globulus* populations. *Canadian Journal of Forest Research* **32**, 59-66.

Jurka J, Pethiyagoda C (1995) Simple repetitive DNA sequences from primates: compilation and analysis. *Journal of Molecular Evolution* **40**, 120-126.

Kalinowski ST (2002) Evolutionary and statistical properties of three genetic distances. *Molecular Ecology* **11**, 1263-1273.

Kalisz S, Nason JD, Hanzawa FM, Tonsor SJ (2001) Spatial population genetic structure in *Trillium grandiflorum*: the roles of dispersal, mating, history, and selection. *Evolution* **55**, 1560-1568.

Keil M, Griffin AR (1994) Use of random amplified polymorphic DNA (RAPD) markers in the discrimination and verification of genotypes in *Eucalyptus*. *Theoretical and Applied Genetics* **89**, 442-450.

Kimura M, Ohta T (1971) 'Theoretical aspects of population genetics.' (Princeton: Princeton)

Kinloch BB (1992) Distribution and frequency of a gene for resistance to white pine blister rust in natural populations of sugar pine. *Canadian Journal of Forest Research* **70**, 1319-1323.

Knowles P (1985) Comparison of isozyme vartiation among natural stands and plantations: jack pine and black spruce. *Canadian Journal of Forest Research* **15**, 902-908.

Knowles P, Perry DJ, Foster HA (1992) Spatial genetic-structure in 2 tamarack [*Larix-Laricina* (Du-Roi) K-Koch] populations with differing establishment histories. *Evolution* **46**, 572-576.

Kremer A (1994) Diversité génétique et variabilité des caractères phénotypiques chez les arbres forestiers. *Genetics, Selection, Evolution* **26**, 105-123.

Kyle CJ, Davis CS, Strobeck C (2000) Microsatellite analysis of North American pine marten (*Martes americana*) populations from the Yukon and Northwest Territories. *Canadian Journal of Zoology* **78**, 1150-1157.

Labillardière JJHd (1799) 'Relation du voyage a la recherche de la Pérouse.' (Jensen: Paris)

Latouche-Halle C, Ramboer A, Bandou E, Caron H, Kremer A (2003) Nuclear and chloroplast genetic structure indicate fine-scale spatial dynamics in a neotropical tree population. *Heredity* **91**, 181-190.

Latta RG (2004) Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytologist* **161**, 51-58.

Le Corre V, Roussel G, Zanetto A, Kremer A (1998) Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. III. Patterns of variation identified by geostatistical analyses. *Heredity* **80**, 464-473.

Leberg PL (1992) Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. *Evolution* **46**, 477-494.

Lefèvre F (2004) Human impacts on forest genetic resources in the temperate zone: an updated review. *Forest Ecology and Management* **197**, 257-271.

Lefort F (1999) Microsatellite sequences: a new generation of molecular markers for forest genetics. *Forest Genetics* **6**, 15-20.

Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**, 1659-1673.

Leite SMM, Bonine CA, Mori ES, Do Valle CF, Marino CL (2002) Genetic variability in a breeding population of *Eucalyptus urophylla* S.T. Blake. *Silvae Genetica* **51**, 253-256.

Leite SMM, Bonine, C.A., Lopes, C.R., Mori, E.S., Valle, C.F.do, and Marino, C.L. (1997) Genetic variation within a base population of *Eucalyptus grandis* Hill ex Maiden using RAPD markers. In 'Proceedings of the IUFRO conference on silviculture and improvement of eucalypts'. Salvador, Brazil pp. 71-74

Leonardi S, Menozzi P (1996) Spatial structure of genetic variability in natural stands of *Fagus sylvatica* L. (beech) in Italy. *Heredity* **77**, 359-368.

Levin DA (1986) Breeding structure and genetic variation. In 'Plant ecology'. (Ed. MJ Crawley) pp. 217-250. (Blackwell Scientific Publications: Oxford)

Levin DA (1988) Consequences of stochastic elements in plant migration. *American Naturalist* **132**, 643-651.

Levin DA, Kerster HW (1974) Gene flow in seed plants. *Evolutionary Biology* **7**, 139-218.

Levin SA (1974) Dispersion and population interactions. *The American Naturalist* **108**, 207-227.

Lewis PO, Zaykin D (2001) Genetic data analysis: computer program for the analysis of allelic data (version 1.1). In. pp. Free program distributed by the authors over the internet from: http://lewis.eeb.uconn.edu/lewishome/software.html.

Lewontin RC (1974) 'The genetic basis of evolutionary change.' (Columbia University Press: New York)

Lewontin RC (1985) Population genetics. *Annual Review of Genetics* **19**, 81-102.

Lewontin RC (2002) Directions in evolutionary biology. *Annual Review of Genetics* **36**, 1-18.

Lexer C, Heinze B, Gerber S, Macalka-Kampfer S, Steinkellner H, Ziegenhagen B, Kremer A, Glossl J (2001) Microsatellite analysis of small anonynous seedlot samples from pedunculate oak (*Quercus robur*): a promising approach to monitor the number of different seed parents and pollen donors. *Forest Science* **70**, 239-250.

Li CC, Weeks DE, Chakravarti A (1993) Similarity of DNA fingerprints due to chance and relatedness. *Human Heredity* **43**, 45-52.

Lian C, Miwa M, Hogetsu T (2001) Outcrossing and paternity analysis of *Pinus densiflora* (Japanese red pine) by microsatellite polymorphism. *Heredity* **87**, 88-98.

Linhart YB (1989) Interactions between genetic and ecological patchiness in forest trees and their dependent species. In 'Evolutionary ecology of plants'. (Eds JH Bock and YB Linhart) pp. 393-430. (Westview Press: Boulder, Colorado)

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic-structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* **82**, 1420-1425.

Lucchini V, Galov A, Randi E (2004) Evidence of genetic distinction and long-term population decline in wolves (*Canis lupus*) in the Italian Apennines. *Molecular Ecology* **13**, 523-536.

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics* **152**, 1753-1766.

Lynch M, Walsh B (1998) 'Genetics and analysis of quantitative traits.' (Sinauer Associates, Inc.: Sunderland, MA, U.S.A.)

Maguire TL, Peakall R, Saenger P (2002) Comparative analysis of genetic diversity in the mangrove species *Vicennia marina* (Forsk.) Vierh. (Avicenniaceae) detected by AFLPs and SSRs. *Theoretical & Applied Genetics* **104**, 388-398.

Malaspina P, Cruciani F, *et al.* (1998) Network analyses of Y-chromosomal types in Europe, Northern Africa, and Western Asia reveal specific patterns of geographic distribution. *American Journal of Human Genetics* **63**, 847-860.

Malécot G (1948) 'Les mathématiques de l'hérédité.' (Masson et Cie: Paris)

Manel S, Berthier P, Luikart G (2002) Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. *Conservation Biology* **16**, 650-659.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* **7**, 639-655.

May RM, Endler JA, McMurtrie RE (1975) Gene frequency clines in the presence of selection opposed by gene flow. *The American Naturalist* **109**, 659-676.

McCauley DE, Wade MJ, Breden FJ, Wohltman M (1988) Spatial and temporal variation in group relatedness - Evidence from the imported willow leaf beetle. *Evolution* **42**, 184-192.

McDonald MW, Rawlings M, Butcher PA, Bell JC (2003) Regional divergence and inbreeding in *Eucalyptus cladocalyx* (Myrtaceae). *Australian Journal of Botany* **51**, 393-403.

McGranahan M, Bell JC, Moran GF, Slee M (1997) High genetic divergence between geographic regions in the highly outcrossing species *Acacia aulacocarpa* (Cunn. ex Benth.). *Forest Genetics* **4**, 1-13.

McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution* **17**, 285-291.

McKinnon GE, Jordan GJ, Vaillancourt RE, Steane DA, Potts BM (2004) Glacial refugia and reticulate evolution: the case of the Tasmanian eucalypts. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* **359**, 275-284.

Meagher TR, Thompson E (1986) The relationship between single parent and parent pair genetic likelihoods in geneology reconstruction. *Theoretical & Applied Genetics* **29**, 87-106.

Menozzi P, Piazza A, Cavalli-Sforza LL (1978) Synthetic maps of human gene frequencies in Europeans. *Science* **201**, 786-792.

Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142**, 1061-1064.

Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theoretical and Applied Genetics* **80**, 437-448.

Milligan BG, McMurray CK (1993) Maximum likelihood analysis of male fertility using dominant and codominant genetic markers. *Molecular Ecology* **2**, 275-283.

Moncur MW, Mitchell A, Fripp Y, Kleinschmidt GJ (1995) The role of honey bees (*Apis mellifera*) in eucalypt plantation forestry. *Commonwealth Forestry Review* **74**, 350-354.

Moran GF, Bell JC (1983) *Eucalyptus*. In 'Isozymes in plant genetics and breeding'. (Eds SD Tanksley and TJ Orton.) pp. 423-441. (Elsevier: Amsterdam)

Moran GF, Bell JC (1987) The origin and genetic diversity of *Pinus radiata* in Australia. *Theoretical and Applied Genetics* **73**, 616-622.

Moran GF, Butcher PA, Glaubitz JC (2000) Application of genetic markers in domestication, conservation, and utilisationof genetic resources of Australasian tree species. *Australian Journal of Botany* **48**, 313-320.

Moran GF, Griffin AR (1983) Recent advances in the study of eucalypt breeding systems. *Silvicultura* **31**, 552-555.

Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika* **37**, 17-23.

Murawski DA (1995) Reproductive biology and genetics of tropical trees from a canopy perspective. In 'Forest Canopies'. (Eds MD Lowman and NM Nadkarni) pp. 457-493. (Academic Press: San Diego)

Namkoong G (1997) A gene conservation plan for lobolly pine. *Canadian Journal of Forest Research* **27**, 433-437.

Namkoong G, Kang HC, Brouard JS (1988) 'Tree breeding: principles and strategies.' (Springer-Verlag: New York)

Nei M (1972) Genetic distance between populations. *American Naturalist* **106**, 283-292.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583-890.

Nei M (1987) 'Molecular evolutionary genetics.' (Columbia Press: New York)

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1-10.

Neigel JE (1997) A comparison of alternative strategies for estimating gene flow from genetic markers. *Annual Review of Ecology and Systematics* **28**, 105-128.

Neigel JE (2002) Is Fst obsolete? *Conservation Genetics* **3**, 167-173.

Nesbitt KA, Potts BM, Vaillancourt RE, West AK, Reid JB (1995) Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globulus* (Myrtaceae). *Heredity* **74**, 628-637.

Nybom H (1991) Applications of DNA fingerprinting in plant breeding. In 'DNA fingerprinting: approaches and applications'. (Eds T Burke, G Dolf, J Jeffreys and R Wohlf). (Birkhauser Verlag: Basel)

Oden NL (1984) Assessing the significance of a spatial correlogram. *Geographical Analysis* **16**, 1-16.

Orme RK (1977) *Eucalyptus globulus* provenances. Canberra, Australia pp. 207-222. (CSIRO, Canberra)

Ouborg NJ, Piquot Y, Van Groenendael JM (1999) Population genetics, molecular markers and the study of dispersal in plants. *Journal of Ecology* **87**, 551-568.

Paetkau D, Waits LP, Clarkson PL, Craighead L, Vyse E, Ward R, Strobeck C (1998) Variation in genetic diversity across the range of North American brown bears. *Conservation Biology* **12**, 418-429.

Parker HG, Kim LV, *et al.* (2004) Genetic structure of the purebred domestic dog. *Science* **304**, 1160-1164.

Parker KC, Hamrick JL, Parker AJ, Nason JD (2001) Fine-scale genetic structure in *Pinus clausa* (Pinaceae) populations: effects of disturbance history. *Heredity* **87**, 99-113.

Parker PG (1998) What molecules can tell us about populations: choosing and using a molecular marker. *Ecology* **79**, 361-382.

Paterson AH (2002) What has QTL mapping taught us about plant domestication? *New Phytol* **154**, 591-608.

Paterson AH, Lin YR, Li Z, Schertz KF, Doebley JF, Pinson S, Liu S, Stansel J, Irvine J (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* **269**, 1714-1718.

Patterson B, Vaillancourt RE, Pilbeam DJ, Potts BM (2004) Factors affecting variation in outcrossing rate in *Eucalyptus globulus*. *Australian Journal of Botany* **52**, 773-780.

Patterson B, Vaillancourt RE, Potts BM (2001) Eucalypt seed collectors: beware of sampling seed lots from low in the canopy! *Australian Forestry* **64**, 139-142.

Patterson B, Wolbang CM, Vaillancourt RE, Potts BM (2000) Inheritance of two chlorophyll mutants in *Eucalyptus globulus*. *Silvae Genetica* **49**, 290-292.

Pavesi A (2004) Detecting traces of prehistoric human migrations by geographic synthetic maps of polyomavirus JC. *Journal of Molecular Evolution* **58**, 304-313.

Peakall R, Smouse PE (2001) GenAlEx V5: Genetic Analysis in Excel. Population genetic software for teaching and research. In. pp. http://www.anu.edu.au/BoZo/GenAlEx. (Australian National University, Canberra, Australia)

Peakall R, Smouse PE, Huff DR (1995) Evolutionary implications of allozyme and rapid variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. *Molecular Ecology* **4**, 135-147.

Pearse DE, Crandall KA (2004) Beyond Fst: Analysis of population genetic data for conservation. *Conservation Genetics* **5**, 585-602.

Peters GB, Lonie JS, Moran GF (1990) The breeding system, genetic diversity and pollen sterility in *Eucalyptus pulverulenta*, a rare species with small disjunct populations. *Australian Journal of Botany* **38**, 559-70.

Pflieger Sp, Lefebvre Vr, Causse M (2001) The candidate gene approach in plant genetics: a review. *Molecular Breeding* **7**, 275-291.

Piazza A, Menozzi P, Cavalli-Sforza LL (1981) The making and testing of geographic gene-frequency maps. *Biometrics* **37**, 635-659.

Poly WJ (1997) Nongenetic variation, genetic-environmental interations and altered gene expression. I- Temperature, photoperiod, diet, pH and sex related effects. *Comparative Biochemistry and Physiology. Part A- Molecular & Integrative Physiology* **117**, 11-66.

Poncet V, Lamy F, Devos KM, Gale MD, Sarr A, Robert T (2000) Genetic control of domestication traits in pearl millet (*Pennisetum glaucum* L., Poaceae). *Theoretical & Applied Genetics* **100**, 147-159.

Potts BM (1990) The response of eucalypt populations to a changing environment. *Tasforests* **2**, 179-193.

Potts BM, Reid. JB (1988) Hybridisation as a dispersal mechanism. *Evolution* **42**, 1245-1255.

Potts BM, Vaillancourt RE, *et al.* (2004) Exploration of the *Eucalyptus globulus* gene pool. In 'Eucalyptus in a changing world. Proceedings of IUFRO Conference'. Aveiro, 11-15 Oct. 2004. (Eds NMG Borralho, JS Periera, C Marques, J Coutinho, M Madeira and M Tomé) pp. 46-61. (RAIZ, Instituto Investigação de Floresta e Papel, Portugal)

Potts BM, Wiltshire RJE (1997) Eucalypt genetics and genecology. In 'Eucalypt ecology: individuals to ecosystems'. (Eds J Wiliams and J Woinarski). (Cambridge University Press: Cambridge)

Pound LM, Patterson B, Wallwork MAB, Potts BM, Sedgley M (2003) Pollen competition does not affect the success of self-pollination in *Eucalyptus globulus* (Myrtaceae). *Australian Journal of Botany* **51**, 189-195.

Pound LM, Wallwork MAB, Potts BM, Sedgley M (2002) Self-incompatibility in *Eucalyptus globulus* ssp. *globulus* (Myrtaceae). *Australian Journal of Botany* **50**, 365-372.

Pritchard J, Falush D, Stephens M (2002) Inference of population structure in recently admixed populations. *American Journal of Human Genetics* **71**, 177-177.

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. *American Journal of Human Genetics* **65**, 220-228.

Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. *American Journal of Human Genetics* **67**, 170-181.

Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution* **16**, 142-147.

Pryor LD (1976) 'The Biology of eucalypts.' (Edward Arnold: London)

Purdie RW (1977a) Early stages of regeneration after burning in dry Sclerophyll vegetation by seed germination. In pp. 34-46)

Purdie RW (1977b) Early stages of regeneration after burning in dry Sclerophyll vegetation. I- Regeneration of the understorey by vegetative means. In pp. 20-34)

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic-markers. *Evolution* **43**, 258-275.

Queller DC, Strassmann JE, Hughes CR (1993) Microsatellites and kinship. *Trends in Ecology & Evolution* **8**, 285-288.

Rajora OP (1999) Genetic biodiversity impacts of silvicultural pracitices and phenotypic selection in white spruce. *Theoretical and Applied Genetics* **99**, 954-961.

Ralph J, J.J. M, Hatfield RD, O'Mally DM, Whetten RW, Sederoff RR (1997) Abnormal lignin in a loblolly pine mutant. *Science* **277**, 235-239.

Rao CR (1973) 'Linear Statistical Inference and its Applications 2nd edition.' (New York)

Reid JB, Hill RS, Brown MJ, Hovenden MJ (Eds) (1999) 'Vegetation of Tasmania.' Flora of Australia Supplementary Series No. 8 (Australian Biological Resources Study: Canberra)

Reif JC, Zhang P, Dreisigacker S, Warburton ML, van Ginkel M, Hoisington D, Bohn M, Melchinger AE (2005) Wheat genetic diversity trends during domestication and breeding. *Theoretical & Applied Genetics* **February GET**, GET.

Rendine S, Piazza A, Menozzi P, Cavalli-Sforza LL (1999) A problem with synthetic maps: reply to Sokal *et al. Human Biology* **71**, 15-25.

Rice WR (1989) Analysing tables of statistical tests. *Evolution* **43**, 223-225.

Rieseberg LH, Burke JM (2001) The biological reality of species: gene flow, selection, and collective evolution [Review]. *Taxon* **50**, 47-67.

Ritland K (1989) Gene diversity and the genetic demography of plant populations. In 'Plant population genetics, breeding, and genetic resources'. (Eds AHD Brown, MT Clegg, AL Kahler and BS Weir). (Sinauer: Sunderland)

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research* **67**, 175-185.

Ritland K (2002) Extensions of models for the estimation of mating systems using n independent loci. *Heredity* **88**, 221-228.

Robledo-Arnuncio JJ, Alia R, Gil L (2004) Increased selfing and correlated paternity in a small population of a predominantly outcrossing conifer, *Pinus sylvestris*. *Molecular Ecology* **13**, 2567-2577.

Rosenberg NA, Burke T, *et al.* (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* **159**, 699-713.

Rosenberg NA, Li L, Ward R, Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. *American Journal of Human Genetics* **73**, 620-620.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. *Science* **298**, 2381-2385.

Rossetto M, Slade RW, Baverstock PR, Henry RJ, Lee LS (1999) Microsatellite variation and assessment of genetic structure in tea tree (*Melaleuca alternifolia* Myrtaceae). *Molecular Ecology* **8**, 633-643.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**, 58-62.

Ruiz A, Barbadilla A (1995) The contribution of quantitative trait loci and neutral marker loci to the genetic variances and covariances among quantitative traits in random mating populations. *Genetics* **139**, 445-455.

Russell J, Marshall D, Griffin D, Harbard J, Powell W (2001) Gene flow in South American *Eucalyptus grandis* and *E. globulus* seed orchards. In 'Developing the eucalypt of the future. Proceedings of IUFRO International Symposium'. Valdivia, Chile

Ryman N, Jorde PE (2001) Statistical power when testing for genetic differentiation. *Molecular Ecology* **10**, 2361-2373.

Sale M, Potts BM, West A, Reid J (1996) Molecular differentiation within and between *Eucalyptus risdonii*, *E. amygdalina* and their hybrids using RAPD markers. *Australian Journal of Botany* **44**, 559-569.

Sampson JF (1998) Multiple paternity in *Eucalyptus rameliana* (Myrtaceae). *Heredity* **81**, 349-355.

Sampson JF, Hopper SD, James SH (1989) The mating system and population genetic structure in a bird-pollinated mallee, *Eucalyptus rhodantha*. *Heredity* **63**, 383-393.

Sanhueza R, Griffin AR (2001) FAMASA Fiber Yield Improvement Programme (F.Y.I.P.): 10 years experience breeding *Eucalyptus globulus*. In 'Developing the eucalypt of the future. IUFRO International Symposium'. (Ed. S Barros). (INFOR, Chile)

Savva M, Potts BM, Reid JB (1988) The breeding system and gene flow in *Eucalyptus urnigera*. In 'Pollination 88'. (Eds RB Knox, MB Sing and LF Troiani) pp. 176-182. (Plant Cell Biology Research Centre: University of Melbourne: Melbourne)

Schmidtling RC, Hipkins V (1998) Genetic diversity in longleaf pine (*Pinus palustrisl*): influence of historical and prehistorical events. *Canadian Journal of Forest Research* **28**, 1135-1145.

Schnabel A, Beerli P, Estoup A, Hillis D (1998) A guide to software packages for data analysis in molecular ecology. In 'Advances in molecular ecology'. (Ed. Carvalho) pp. 291-303. (IOS Press: Oxford)

Schnabel A, Hamrick JL (1995) Understanding the population genetic-structure of *Gleditsia triacanthos* L. - The scale and pattern of pollen gene flow. *Evolution* **49**, 921-931.

Shepherd M, Jones M (2005) Molecular markers in tree improvement: characterisation and use in *Eucalyptus*. In 'Molecular marker systems in plant breeding and crop improvement'. (Eds H LORZ and G WENZEL) pp. 399-409. (Springer-Verlag: Heidelberg)

Shi GR (1993) Multivariate data-analysis in paleoecology and paleobiogeography - A review. *Palaeogeography Palaeoclimatology Palaeoecology* **105**, 199-234.

Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R (1995) A novel measure of genetic-distance for highly polymorphic tandem repeat loci. *Molecular Biology and Evolution* **12**, 914-920.

Skabo S, Vaillancourt RE, Potts BM (1998) Fine-scale genetic structure of *Eucalytus globulus* ssp. *globulus* forest revealed by RAPDs. *Australian Journal of Botany* **46**, 583-594.

Slatkin M (1973) Gene flow and selection in a cline. *Genetics* **75**, 733-756.

Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**, 393-430.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**, 264-79.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 1463-1463.

Slatkin M, Arter HE (1991) Spatial autocorrelation methods in population-genetics. *American Naturalist* **138**, 499-517.

Smith S, Hughes J, Wardell-Johnson G (2002) High population differentiation and extensive clonality in a rare mallee eucalypt: *E. curtisii*. *Conservation Genetics* **00**, 1-12.

Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution* **55**, 260-271.

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extinsions of the Mantel test of matrix correspondence. *Systematic Zoology* **35**, 627-632.

Smouse PE, Meagher TR, Kobak CJ (1999) Parentage analysis in *Chamaelirium luteum* (L.) Gray (Liliaceae): why do some males have higher reproductive contributions. *Journal of Evolutionary Biology* **12**, 1069-1077.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* **82**, 561-573.

Smouse PE, Sork VL (2004) Measuring pollen flow in forest trees: an exposition of alternative approaches. *Forest Ecology and Management* **197**, 21-38.

Sokal RR, Jacquez GM, Wooten MC (1989) Spatial auto-correlation analysis of migration and selection. *Genetics* **121**, 845-855.

Sokal RR, Oden NL (1977) Spatial autocorrelation in biology. 2- Some biological implications and four applications of evolutionary and ecological interest. *Biological Journal of the Linnean Society* **10**, 229-249.

Sokal RR, Oden NL (1978) Spatial autocorrelation in biology 1. Methodology. *Biological Journal of the Linnean Society* **10**, 199-228.

Sokal RR, Oden NL, Thomson BA (1997) A simulation study of microevolutionary inferences by spatial autocorrelation analysis. *Biological Journal of the Linnean Society* **60**, 73-93.

Sokal RR, Rohlf FL (1981) 'Biometry.' (W. H. Freeman: San Francisco)

Soria F, Basurco F, Toval G, Silio L, Rodriguez MC, Toro M (1998) An application of Bayesian techniques to the genetic evaluation of growth traits in *Eucalyptus globulus*. *Canadian Journal of Forest Research* **28**, 1286-1294.

Sork VL (1998) Introduction to workshop: studying gene flow on an ecological time scale. In 'Gene flow in fragmented, managed and continuous populations'. National Center for Ecological Analysis and Synthesis, University of California-Santa Barbara. (Ed. VL Sork)

Sork VL, Nason J, Campbell DR, Fernandez JF (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology & Evolution* **14**, 219-224.

Steane DA, Vaillancourt RE, Russell J, Powell W, Marshall D, Potts BM (2000) Microsatellite variation in *Eucalyptus globulus* (Myrtaceae). *Molecular Ecology*, submitted.

Steane DA, Vaillancourt RE, Russell J, Powell W, Marshall D, Potts BM (2001) Microsatellite variation in a forest tree, *Eucalyptus globulus* ssp. *globulus* (Myrtaceae). *Silvae Genetica* **50**, 89-91.

Stoehr MU, Elkassaby YA (1997) Levels of genetic diversity at different stages of the domestication cycle of interior spruce in British Columbia. *Theoretical & Applied Genetics* **94**, 83-90.

Streiff R, Ducousso A, Lexer C, Steinkellner H, Gloessl J, Kremer A (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L-and *Q-petraea* (Matt.) Liebl. *Molecular Ecology* **8**, 831-841.

Streiff R, Labbe T, Bacilieri R, Steinkellner H, Glossl J, Kremer A (1998) Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozmes and microsatellites. *Molecular Ecology* **7**, 317-328.

Sunnocks P (2000) Efficient genetic markers for population biology. *Trends in Ecology & Evolution* **15**, 199-203.

Szmidt AE, Muona O (1985) Genetic effects of Scots pine (*Pinus sylvestrus* L.) domestication. *Lecture Notes in Bioinformatics* **60**, 241-252.

Tonsor SJ, Kalisz S, Fisher J, Holtsford TP (1993) A life-history based study of population genetic-structure - Seed bank to adults in *Plantago lanceolata. Evolution* **47**, 833-843.

Toro MA, Silió L, Rodriguez MC, Soria F, Toval G (1998) Genetic analysis of survival to drought in *Eucalyptus globulus* in Spain. In '6th World Congress on Genetics Applied to Livestock Production'. Armidale, Australia pp. 499-502. (Organising Committee)

Trustrum GB (1961) The correlations between relatives in a random mating diploid population. *Proceedings of the Cambridge Philosophical Society* **57**, 315-312.

Turner ME, Stephens JC, Anderson WW (1982) Homozygosity and patch structure in plant populations as a result of nearest-neighbor pollination. *PNAS* **79**, 203-207.

Ueno S, Tomaru N, Yoshimaru H, Manabe T, Yamamoto S (2000) Genetic structure of *Camellia japonica* L. inan old-growth evergreen forest, Tsushima, Japan. *Molecular Ecology* **9**, 647-656.

Vaillancourt RE, Potts BM, Watson M, Volker PW, Hodge GR, Reid JB, West AK (1995) Detection and prediction of heterosis in *Eucalyptus globulus. Forest Genetics* **2**, 11-19.

Vaillancourt RE, Skabo S, Gore PL (1998) Fingerprinting for quality control in breeding and deployment. *Australian Forestry* **61**, 207-210.

Van de Casteele T, Galbusera P, Matthysen E (2001) A comparison of microsatellite-based pairwise relatedness estimators. *Molecular Ecology* **10**, 1539-1549.

Vazquez-Dominguez E, Paetkau D, Tucker N, Hinten G, Moritz C (2001) Resolution of natural groups using iterative assignment tests: an example from two species of Australian native rats (Rattus). *Molecular Ecology* **10**, 2069-2078.

Vega Alonso G, Toval Hernandez G, Basurco Garcia-Casal F, Rodriguez Soalleiro R (1994) Early growth of 260 families of *Eucalyptus globulus* in North Spain. In '*Eucalyptus* for biomass production'. Lisbon, Portugal. (Eds J Pereira and H Pereira) pp. 115-132. (Commission of the European Communities)

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* **13**, 921-935.

Verrier E, Colleau JJ, Foulley JL (1994) Genetic variablity when selecting on the animal model BLUP. In 'Proceedings of the 5th World Congress on Genetics Applied to Livestock Production' pp. 139-142. (University of Guelph)

Vignal A, Milan D, SanCristobel M, Eggen A (2002) A review of SNP and other types of molecular markers and their use in animal genetics. *Genetics, Selection, Evolution* **34**, 275-305.

Wade MJ, Mccauley DE (1988) Extinction and recolonization - Their effects on the genetic differentiation of local populations. *Evolution* **42**, 995-1005.

Wang JL (2002) An estimator for pairwise relatedness using molecular markers. *Genetics* **160**, 1203-1215.

Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. *Nature* **398**, 236-238.

Wang XR, Szmidt AE (2001) Molecular markers in population genetics of forest trees. *Scandinavian Journal of Forest Research* **16**, 199-220.

Waser NM, Price MV (1983) Optimal and actual outcrossing in plants and the nature of plant-pollinator interaction. In 'Handbook of experimental pollination biology'. (Eds CE Jones and RJ Little) pp. 341-359. (Van Nostrand Reinhold Co. Inc.: New York)

Weir BS (1996) 'Genetic data analysis II: Methods for discrete population genetic data.' (Sinauer: Sunderland)

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.

Weir BS, Hill WG (2002) Estimating F-statistics. *Annual Review of Genetics* **36**, 721-750.

Wettenhall R (2005) Bushfires. In 'Biology and Biologists entry in The Companion to Tasmanian
History'. (Ed. A Alexander) pp. 59-60. (Centre for Tasmanian Historical Studies, School of
History and Classics, University of Tasmania: Hobart)

Wickneswari R, Norwati M (1993) Genetic diversity of natural populations of *Acacia auriculiformis*. *Australian Journal of Botany* **41**, 65-77.

Williams CG, Hamrick JL (1995) Genetic diversity levels in an advanced generation *Pinus taeda* L. program measured using molecular markers. *Forest Genetic Resources Newsletter* **23**, 45-50.

Williams K, Potts BM (1996) The natural distribution of *Eucalyptus* species in Tasmania. *Tasforests* **8**, 39-164.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177-1191.

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *PNAS* **84**, 9054-9058.

Wolfe KH, Sharp PM, Li WH (1989) Mutation rates differ among regions of the mammalian genome. *Nature* **337**, 283-285.

Wright S (1931) Evolution in Medelian populations. *Genetics* **16**, 290-s-.

Wright S (1943) Isolation by distance. *Genetics* **28**, 115-136.

Wright S (1948) On the role of directed and random changes in gene frequency in the genetics of populations. *Evolution* **2**, 279-294.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics* **15**, 323-354.

Xiong LZ, Liu KD, Dai XK, Xu CG, Zhang Q (1999) Identification of genetic factors controlling domestication-related traits of rice using an F2 population of a cross between *Oryza sativa* and *O. rufipogon*. *Theoretical & Applied Genetics* **98**, 243-251.

Yanchuk AD (2001) A quantitative framework for breeding and conservation of forest tree genetic resources in British Columbia. *Canadian Journal of Forest Research* **31**, 566-576.

Yeh FC, El-Kassaby YA (1980) Enzyme variation in natural populations of Sitka spruce (*Picea sitchensis*). I. Genetic variation patterns among trees from 10 IUFRO provenances. *Canadian Journal of Forest Research* **10**, 415-422.

Young AG, Merriam HG (1994) Effects of forest fragmentation on the spatial genetic-structure of *Acer saccharum* marsh (sugar maple) populations. *Heredity* **72**, 201-208.

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology* **11**, 1-16.

Zang D, Wang H, You Y (1995) Performance and selection of a 4-year *Eucalyptus globulus* seedling seed orchard in Yunnan, China. In 'Eucalypt plantations: improving fibre yield and quality'. (Eds BM Potts, NMG Borralho, JB Reid, RN Cromer, WN Tibbits and CA Raymond) pp. 226-229. (CRC for Temperate Hardwood Forestry, Hobart, Tasmania)

Zhang D, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* **12**, 563-584.

Zheng YQ, Ennos RA (1999) Genetic variability and structure of natural and domesticated populations of Caribbean pine (Pinus caribaea Morelet). *Theoretical & Applied Genetics* **98**, 765-771.